

Africa 2005

The world's poorest continent is rightly at the top of the global agenda this year. But the agenda needs to be set by Africa, with the outside world in a supporting role — not the other way round.

Tackling the neglect of Africa is high on the list of priorities of the G8 — the world's eight largest industrialized nations — and the European Union (EU) in 2005. This neglect was once described as a “scar on the conscience of the world” by Tony Blair, Britain's prime minister, who this year chairs the G8. Chief among the items for action are a promise of debt relief for the poorest countries, improved terms of trade with the EU, and up to \$50 billion each year in aid through a planned fund known as the International Finance Facility.

It is likely that Blair and the UK chancellor, Gordon Brown, will succeed in getting endorsement for their plans from many G8 leaders. But that still leaves two bigger obstacles: raising the money, and deciding how to spend it. In a report published in January, Jeffrey Sachs, an adviser to the United Nations secretary-general Kofi Annan, said that to meet the Millennium Development Goals — a series of targets to halve global poverty, hunger and disease — rich countries will need to raise their annual aid in support of the goals by a factor of four to \$121 billion in 2006, rising still further to \$189 billion in 2015.

If additional aid cheques begin to flow, who will decide how the money is spent? To help Africans set their priorities, Blair and Bob Geldof set up the Commission for Africa — a group of 17 commissioners comprising world leaders and heads of UN agencies (mostly from Africa) who have been charged with asking people what they think international aid should be spent on.

Technology transfer

Over the past few months, the commission has organized meetings and online discussion groups, canvassing thousands of people, including politicians, professionals, the business community and non-governmental organizations in many countries. The results of this innovative exercise will be published in a report in March and handed to the G8 leaders when they meet in Scotland in July.

In one of the commission's interim reports seen by *Nature*, there is strong support for more international aid to be spent on strengthening science and technology in Africa, including the transfer of Western technology to the continent, as well as accelerated efforts towards a malaria vaccine. Some of these findings seem to have taken the commissioners by surprise: they didn't seem to think (at least initially) that Africans would consider boosting science and technology to be a priority.

To their credit, international aid donors are well aware that there is enthusiasm in Africa for more aid to be spent on science and technology. Of all the developing countries, those in Africa have found it hardest to absorb and adapt technology from abroad, and have lagged behind in setting up their own research institutions in sufficiently large numbers to create a critical mass of productive researchers. According to the Sachs report, African countries have on average only 18 scientists and engineers per million people, compared with 69 in southern Asia, 273 in Latin America, and 903 in eastern Asia.

Substantive action has already been taken. The Bill and Melinda Gates Foundation — already busy with an ambitious healthcare

research agenda — last week named the science academies of Nigeria, South Africa and Uganda as recipients of a US\$20-million grant. Britain's Department for International Development also has plans to increase its spending on boosting research and development in Africa, and in December it appointed Gordon Conway, former president of The Rockefeller Foundation in New York, to help it devise a new science and technology strategy.

Setting the agenda

All this should provide grounds for hope. Yet it is vital that as aid donors step forward with more money, they are faced with a plan based on meeting real needs on the ground; otherwise they may set the agenda themselves. At a meeting of aid agencies and scientists in London last month, for example, nearly every speaker promised to respond to the needs of Africa. But in reality, the meeting had been called by the Canadian and UK governments to see if they could coordinate their spending priorities. A select group of scientists from Africa were also present, but their status was merely that of invited guests to a party.

In public, few of the African delegates questioned this reversal of roles, perhaps not wanting to offend their hosts and risk cutting off future funding prospects. In private, however, they were more critical. One exception was John Mugabe, a scientific adviser to the New Partnership for Africa's Development (NEPAD), based in Pretoria, South Africa. Mugabe said that Africa is changing because Africans want to change, and that change will take place with or without international aid.

Mugabe is right. Multiparty government is at last taking root in Africa. It is early days still, but the leaders of Nigeria, South Africa and Ethiopia are showing a willingness to turn their countries around. African countries see the EU as a model for their own African Union, and under NEPAD they have also agreed to evaluate each other's performance in an effort to expose corruption.

The next generation

In the scientific community, a new generation of able and confident leaders is coming forward, despite the brain drain. Indigenous investment is taking place, too. Nigeria, for example, is in the midst of overhauling its colonial-era research infrastructure with the help of Japan and UNESCO. Kenya is hosting several initiatives to boost agricultural research. Mozambique and Tanzania are showing that centres of excellence in healthcare research can be set up and run in a resource-scarce environment.

Africa is clearly on the move. But so far it seems that aid donors in rich countries are a little confused as to how they should respond. This may be partly because, as they admit, they are historically used to telling people in Africa what is best for them, and then complaining when their plans go awry. Aid agencies are now encountering a different generation of Africans and must learn to change gear. They should empower many more Africans with the confidence to chart their own futures. And they must have the confidence themselves to trust the judgement of those who already have a clear idea of what they want. ■





Welcome pack
US immigration improves rules for visiting scientists
p672



Stock answer
Digital link helps customs officers identify bugs
p673



Power struggle
Flooded reactor reignites India's energy debate
p675



Weight lifter
Europe's updated Ariane rocket blasts off successfully
p676

Salt sellers challenge US health agency using data-quality act

Meredith Wadman, Washington

Health agencies in the United States are being leant on by the private sector to release the raw data that underpin disputed government regulations.

The pressure is being brought to bear through the Data Quality Act, a law championed by the Bush administration. The act allows companies, activist groups and citizens to challenge government statements and rules, prompting corrections if the government cannot demonstrate their scientific validity.

The measure worried scientists when it took effect in late 2002 (see *Nature* **416**, 249; 2002). Its effect is now being felt at several research agencies — including the National Institutes of Health (NIH). As early as this summer, for example, a US Court of Appeals will judge a plea from the Virginia-based Salt Institute, which represents salt producers. The institute wants direct access to the data behind a study that linked salt consumption to high blood pressure.

The trial, funded by the National Heart, Lung, and Blood Institute (NHLBI), studied the impact of dietary sodium intake on blood pressure and the results were published in *The New England Journal of Medicine*, *Annals of Internal Medicine* and *The American Journal of Cardiology*. They showed that reducing dietary sodium lowers blood pressure in most people, and this led the government to recommend that Americans consume less salt.

Researchers in the trial say that they have released all the data the Salt Institute could want or need — and that it is misusing the act. “It is trying to slice and dice the data set so it finds a group that seems not to have a blood pressure that’s responsive to reduction in salt,” says Lawrence Appel, a physician at Johns Hopkins School of Public Health in Baltimore, Maryland, and one of the trial’s principal investigators. “That’s blatantly inconsistent with a scientific approach to analysing clinical data.”

Last month, the Salt Institute and the US



Spoon-fed: the Salt Institute is fighting for access to unpublished data.

Chamber of Commerce asked the Fourth Circuit Court of Appeals to overturn a decision by a lower court in Virginia. That court had ruled that the NHLBI was within its rights in refusing to release the data that had been requested.

“If any agency deserves to be sued, here’s one,” says Richard Hanneman, the president of the Salt Institute. He says that the NHLBI has shown a consistent “pattern of obfuscation and non-responsiveness” over requests for access to unpublished data, such as the initial blood pressures of the trial’s 412 subjects.

The NHLBI declined to comment while the matter is before the courts.

Rising pressure

The salt case is the latest in a flurry of data-quality actions against NIH institutes. Last year, for example, the National Institute on Aging rewrote a statement on its website, which claimed that chewing tobacco is as bad for health as smoking, after a private think-tank, funded by conservative activist Richard Scaife, requested a correction. And Jim Tozzi, the industry-affiliated lobbyist who actually wrote most of the law, has

challenged the chemical review procedures of the NIH-run National Toxicology Program and its investigation of possible carcinogens.

However, the salt case marks the first time that a petitioner has actually sued under the Data Quality Act. In other cases, complainants have simply petitioned the relevant government agencies. The *Washington Post* reported last summer that, since the law came into force, industry has filed four-fifths of the substantive petitions under the law; environmental or citizen groups have filed the others.

The petitions have targeted bodies such as the Environmental Protection Agency and the Fish and Wildlife Service. Last month, the Fish and Wildlife Service decided not to list the sage grouse as an endangered species, against environmentalists’ advice, after an

industry group and a county board in Idaho challenged the data that the agency was using to support the listing.

Critics of the act say that it was devised primarily as a tool for conservatives to attack unwelcome regulations by picking at isolated pieces of the science behind them, rather than weighing all the evidence and coming to appropriate conclusions.

“The effort is to suppress the science, to destroy its credibility even though it’s widely available and has been vetted many times,” says Rena Steinzor, a law professor at the University of Maryland in Baltimore.

The law’s backers sharply disagree. Tozzi says that when federal policies have effects that cost millions or billions of dollars, the public should have an opportunity to challenge the science behind them. Now an adviser at the Center for Regulatory Effectiveness, a Washington-based government watchdog, Tozzi says that the great majority of scientists should be unaffected by the act. But if scientists “are going to get in the regulatory sandbox, they gotta play by these rules”, he says. “They can’t use their white coats to shield them from the sunlight.”

US extends security clearance for scholars

Geoff Brumfiel, Washington

The US government has announced that visa-related security checks will remain valid for up to four years. The planned change will ease travel in and out of the country, and greatly reduce the chance that students or scientists are left stranded abroad.

Academic groups are delighted with the move, which they say will simplify visitors' trips home. "It's fantastic," says Wendy White, head of the international office at the National Academy of Sciences in Washington. "I think it will make a big difference for visiting students and scholars."

The Department of State and the Department of Homeland Security, which are jointly responsible for immigration policy, revealed their decision on 11 February. "This change sends a clear message that the United States encourages those with great scientific minds to explore studying and working in our country," said a statement from Asa Hutchinson, under-secretary for borders and transportation at the homeland security department.

Since the terrorist attacks of 11 September 2001, background checks have been



Hail fellow: visa problems have dogged scientists visiting the United States.

required for many scientists seeking visas to work or study in the United States. In the past, the checks have led to lengthy delays that caused some researchers to miss meetings or the start of academic years. Even students who had studied in the United States for years found themselves unable to return until their security checks had been completed (see *Nature* 427, 190–195; 2004).

Under the new scheme, security clearances for students will be valid for four years, and those for scientists working in the United States on scholar or work visas will last for two years. In addition, researchers

visiting the country for conferences and other events will hold valid security clearance for 12 months. The net effect is that researchers will be able to move in and out of the United States more easily after an initial check has been done, according to Angela Aggeler, a spokeswoman for the state department.

Aggeler adds that the average time taken to make a security check has already been slashed from 75 to 14 days over the past year. "We will continue to try and lower the waiting time," she says.

"I think that this kind of change sends a good signal to the international community," says Heath Brown, director of research and policy analysis at the Council of Graduate Schools in Washington. A survey by the council last year showed that the number of foreign students admitted to US schools was down in 2004 (see *Nature* 431, 231; 2004). Brown says that he hopes the changes will bring foreign students and scholars back to US universities, but he points out that many still see the country as an unwelcoming place in the wake of 9/11. White agrees: "We have to get the word out that things really are better."

Nuclear-physics research falls foul of budget cuts

Jessica Ebert, Washington

US physicists are hitting out hard at the Bush administration's budget proposal for next year. They claim that the plan, which would cut overall spending on research in physics by about 4%, will imperil the country's global leadership in the discipline.

"If these cuts stand," says Tom Ludlam, a physicist at Brookhaven National Laboratory in New York state, "there will be a significant reduction in the number of PhD scientists, graduate students, postdocs and senior scientists doing physics here in America."

The Department of Energy's Office of Science, which funds most US physics research, formulated a plan in November 2003 to construct or upgrade 28 different facilities over the next 20 years. Physicists say that the proposed budget will prevent the plan from being implemented.

If approved by Congress, the budget request would support some of the facilities, including construction of the Linac Coherent Light Source at the Stanford Linear Accelerator Center in California. But it

would cancel or delay several other projects that the plan identified as priorities.

For instance, the budget would cancel the BTeV experiment, which aims to use the Tevatron collider at Fermilab, Chicago, to investigate the imbalance between matter and antimatter in the Universe. The experiment received a top ranking in the 2003 plan. "This is quite a blow to us," says project director Joel Butler. "Everything was in place," he adds. "We were more than ready to go."

The science-office budget has been essentially flat since 2001, but Raymond Orbach, head of the energy department's science office, defends the 2006 proposal. "I'm very happy with the structure of the budget," he says. "It enables us to maintain scientific leadership on a global scale."

Martha Krebs, a consultant in Los Angeles who ran the science office under Bill Clinton, says the budget reveals the areas in which the Bush administration wants to "be a leader". She says that nuclear physics, in particular, does not seem to be one of these.

The 2006 nuclear-physics request for US\$371 million is 8% less than this year's budget. Brookhaven and the Thomas Jefferson National Accelerator Facility in Newport News, Virginia, would bear the brunt of the cut. At the Jefferson lab, a plan to double the energy of the laboratory's continuous electron beam — also ranked highly in the 2003 plan — would be deferred. In addition, the electron beam's operating hours would be cut by almost one-third.

At Brookhaven, the cut would decrease operating hours at the Relativistic Heavy Ion Collider — the laboratory's main nuclear-physics facility — by almost two-thirds. "It's not a good situation," says Rick Casten, a physicist at Yale University and chairman of the Nuclear Science Advisory Committee. "We're going to lose our competitiveness if this keeps up and is not reversed."

For more news and analysis go to
news @ nature.com
www.nature.com/news



Pin-ups: researchers will soon be able to view 28 million insects through a microscope on the Internet.

Online access offers fresh scope for bug identification

Jim Giles, London

When Michael Schauff's work hits a snag, millions of dollars can be at risk. Insects contaminate thousands of US agricultural imports every year, and Schauff's lab in Beltsville, Maryland, is charged with identifying the immigrant critters. If his team lacks a reference specimen needed for the identification, decisions about whether to quarantine the cargo can become difficult.

Help may soon be at hand. Thanks to a remote-control microscope currently under development, Schauff will soon be able to access another 28 million specimens — those in the entomology collection of London's Natural History Museum (NHM). And advocates of the system suggest that it's about far more than customs work — they say the link-up shows how emerging digital technology will transform taxonomy.

When the microscope is installed this May, the US researchers will simply phone through requests for specimens. After a technician in London puts the insect on the microscope, Schauff's team will be able to use the device almost as if they were in the room with it — adjusting magnification, rotating the specimen and taking high-resolution images. The system is being developed by two Virginia-based optics companies, Microptics in Ashland and TriTek in Sumerduck.

"We get 5,000 urgent identifications every year and they need a report back the same day," says Schauff. "That's millions of dollars of stock we're playing with every year."

Quentin Wheeler, head of entomology at the NHM, estimates that the system will cost about US\$135,000 to install at each location. If the technology works, he is keen to expand it to the large insect collections at the natural history museums of Paris and Washington.

Together, the three collections contain more than 100 million specimens covering at least 95% of all described insect species, he says.

Wheeler adds that the technology will help to promote online collaborations between entomologists and, by archiving pictures taken when examining the specimens, it will eventually create a virtual library of insect specimens.

Adding video conferencing to the software used to control the microscopes would allow experts around the world to discuss the specimens, Wheeler says. And a portable version of the microscope would let entomologists guide the fieldwork of collectors.

But some advocates of a new vision for taxonomy question whether the system is the best way forward. "It is what happens next that will be really exciting," says Charles Godfray, director of the Centre for Population Biology in Ascot, near London, pointing out that in the long term, insect collections can be digitized and made permanently available online.

Several databases of two-dimensional images, such as AntWeb, maintained by the California Academy of Sciences, are already up and running. And Godfray says that three-dimensional images can be created by imaging specimens from several angles and using software to stitch the pictures together.

Wheeler agrees that such a database should be created, but adds that researchers will always need to examine the actual specimens. The intricate structure of many insects' bodies will be difficult to capture using a fixed number of images, he suggests. The size of the collections also makes the creation of thorough databases a very long-term goal. "We've got 28 million specimens here," Wheeler says. "You'll have to give us some time."

Europe's research still lacks competitive edge, says panel

Quirin Schiermeier, Munich

The European Union's main research programme has met with only "modest" success in its main goal of strengthening Europe's industrial competitiveness. That's the somewhat sobering conclusion of the latest five-year assessment of the Framework funding programme, published last week.

The report's recommendations echo calls that have come from the science community for several years. These include the establishment of a European Research Council (see *Nature* 425, 440; 2003), the simplification of Framework's notoriously cumbersome funding administration, and the extension of programmes that provide relocation fellowships for young scientists. It also calls for greater participation of small high-tech companies in the next Framework programme, which begins in 2006.

The assessment covers 1999–2003 and was conducted by a 13-strong panel of representatives from both academia and industry, chaired by Erkki Ormala, vice-president of technology policy at the Finnish IT company Nokia.

"Only a few European universities are recognized as global leaders," laments the report. "This is, at least in part, a result of insufficient resources combined with the fragmented nature of the European research-and-technology development landscape." Opening some national science programmes Europe-wide and improving the coordination of national research activities are suggested as ways to strengthen the overall quality of research in the European Union (EU).

Some of the report's recommendations could well be implemented thanks to an anticipated rise in funds for the Framework programme (see *Nature* 433, 96; 2005). The new EU research commissioner, Janez Potočnik, hopes to get approval from the European Parliament this summer for a doubling of the EU's four-year research budget to €30 billion (US\$40 billion).

Scientists in the new EU member states, where national funding opportunities are few, will be watching to see if the budget rise comes through, and if the report's recommendations are followed. "Any proposal that increases our chance of getting funded is very welcome," says Anna Pytko, vice-director of the Polish National Contact Point for EU research in Warsaw. ■
http://europa.eu.int/comm/research/reports/2004/fya_en.html

Scientists urged to end feud with White House

Geoff Brumfiel, Washington

Bill Clinton's former scientific adviser is trying to persuade leading researchers to use the start of President George Bush's second term as an opportunity to patch up their differences with the White House.

Neal Lane, who served under President Clinton from 1998 to 2001, says researchers should meet with new members of Bush's cabinet and try to re-engage with them on issues such as global warming. "When an administration is starting a new term, you've got to do everything you can to help," he says.

Bush's first term was marked by unusually bad relations between scientific leaders and the White House, senior researchers say. Aside from well-publicized disagreements on global warming and stem-cell research, they worry that scientists working inside the government, as well as independent bodies such as the National Academies and the top scientific societies, lack influence within the Bush administration.

Lane was one of more than 150 top scientists to sign a petition in February 2004 that accused the administration of "misrepresenting or suppressing scientific knowledge" (see *Nature* 427, 663; 2004). The statement was accompanied by a report from the Union of Concerned Scientists (UCS), a Boston-based environmental watchdog, which alleged that administration officials had altered or concealed scientific reports that were in conflict with administration policies. In the run up to the election, Lane was one of



Neal Lane believes that researchers can build a better relationship with Bush's new cabinet.

a group of prominent scientists to campaign against Bush on university campuses (see *Nature* 430, 595; 2004).

With the election over, Lane now says that it is time to start afresh with administration officials. He believes that talking to them early on may help avoid the skewing of science for political reasons. "A lot will depend on these new people taking over the agencies," he says. "It seems to me that it wouldn't hurt for Bush administration officials and representatives from the scientific community to meet."

But the election campaign has left scientists and administration officials at logger-

heads, and some observers wonder if this will cause relations between them to deteriorate further. According to Robert Walker, a Washington-based lobbyist and former Republican chairman of the House Committee on Science, administration officials were deeply stung by statements in the UCS report that a senior adviser in the president's Office of Science and Technology Policy lacked adequate scientific training. "In politics, it becomes very difficult to work with people who attack you personally," Walker says.

And scientists in some disciplines, at least, still think their views are being repressed by the administration. According to a UCS survey of scientists working for the US Fish and Wildlife Service released on 9 February, 42% of the 414 respondents think that they cannot raise "concerns about the biological needs of species and habitats without fear of retaliation". Kurt Gottfried, a physicist at Cornell University in Ithaca, New York, and chairman of the UCS, says that the report clearly shows the mistrust between scientists and the Bush administration, and predicts that more surveys being planned by the UCS will give similar results.

Henry Kelly, head of the Washington-based Federation of American Scientists, and an anti-Bush campaigner during the election, also predicts that disputes will continue to rage between scientists and the Bush administration. "This is not a fight that the country can afford," he says. Walker agrees: "I don't think this will fade away," he says. ■

T. LAVERGNE

Sanctions agreed over teenager's gene-therapy death

Erika Check, Washington

Three clinical researchers are to face restrictions on their work for their part in a gene-therapy trial that led to a teenager's death in 1999. An out-of-court settlement, under which their employers will pay fines of \$1 million, was announced on 11 February. It marks the end of a five-year investigation by the US Department of Justice into the death in 1999 of 18-year-old Jesse Gelsinger in the gene-therapy trial.

Under the terms of the settlement, James Wilson, who led the trial at the University of Pennsylvania, is required to undergo retraining in the conduct of trials on human subjects before he can work with them again.

Wilson hasn't worked with human subjects since January 2000 — four months after Gelsinger died in the trial, which used a modified virus to deliver a gene that was intended to produce a liver enzyme that was deficient. According to an investigation by

the university, Gelsinger died from an immune reaction to the adenovirus vector. The widely publicized case led to congressional hearings and to tighter rules on the conduct of clinical trials.

The settlement also places restrictions on the work of Wilson's co-investigators in the gene-therapy trial, Steven Raper, also of the University of Pennsylvania, and Mark Batshaw, of the Children's National Medical Center in Washington. The University of Pennsylvania and the Children's National Medical Center will pay fines totalling \$1 million.

The justice department alleged that the researchers and their institutions made false statements regarding the safety of the trial to the National Institutes of Health, the Food and Drug Administration, and the institutional review board that oversaw the research.

But the researchers and their institutions state in the settlement that they did nothing

wrong, and that their conduct was "at all times lawful and appropriate".

The terms of the settlement state that a monitor will supervise Wilson's work in humans for three years, and he will be allowed to conduct only one trial at a time. Any of Wilson's animal research that could affect patient safety will also be supervised. If he fulfils the requirements set out in the settlement, Wilson will then be able to conduct unrestricted research in humans in 2010.

In a statement, Wilson said that the settlement would enable him to continue with his laboratory research. But Jesse Gelsinger's father, Paul Gelsinger, says he is disappointed that the settlement did not require Wilson or anyone else to admit responsibility, apologize for their actions, or release documents uncovered by the justice department. "We just want that kind of closure," he says. "Without that, I'm finding it impossible to forgive these guys." ■



Safety first: the fast-breeder reactor at Kalpakkam, India, was submerged by December's tidal waves.

India's nuclear debate hots up after tsunami floods reactor

K. S. Jayaraman, New Delhi

India's nuclear regulator has called for a detailed report on the impact of December's tsunami on the nation's prototype fast-breeder reactor. The reactor's site at Kalpakkam in Tamil Nadu state was flooded by the wave, prompting fears about its future safety.

The Atomic Energy Regulatory Board wants to know whether it should modify the design of the Rs35-billion (US\$800-million) project — the centrepiece of India's nuclear-energy research programme — in light of the flood, which killed a construction worker.

Nuclear engineers say the flood shouldn't necessitate any major changes to either the siting or the design of the reactor, but critics are using the flood to reopen a debate about whether the ambitious project is appropriate to India's pressing energy needs.

Fast-breeder reactors were once widely regarded as the future of nuclear power. They rely on chain reactions that produce more fissile material than they consume.

India's prototype reactor at Kalpakkam would use sodium as a coolant and oxides of plutonium and uranium as fuel to produce 500 MW of electricity. It is the first of five such reactors that India hopes to build by 2020, says Anil Kakodkar, secretary of the Department of Atomic Energy, after which it would switch to larger reactors.

But Ashok Parthasarathi, a science-policy specialist at Jawaharlal Nehru University in New Delhi, says the huge project demonstrates the unbalanced nature of energy policy in India. The country spends only about Rs150 million each year on research into renewable energy sources.

V. S. Arunachalam a former scientific adviser to India's defence department who is now at Carnegie Mellon University in Pittsburgh, Philadelphia, says that fast-breeder reactors will not be economically viable until the end of this century. And although France and Russia are still seeking to develop them, the United States, Britain and Germany have each abandoned fast breeders in the face of mounting costs, technical problems and the continued availability of cheap uranium for conventional nuclear reactors.

But Baldev Raj, who heads the Indian project, says the technology "is very much alive" internationally. He adds that India's decision to proceed is based on "experience gained from the design and operation" of a fast-breeder test reactor at Kalpakkam since 1987. Critics counter that the new reactor is 60 times bigger and relies on a different fuel.

Officials at the atomic-energy department say that India has a special need for the technology because it cannot buy uranium from abroad without agreeing to put all its nuclear facilities under International Atomic Energy Agency safeguards — something it has always refused to do.

The Kalpakkam project has been dogged by bad luck so far. India's prime minister, Manmohan Singh, missed a ceremonial pouring of concrete last August because of ill health.

S. K. Sharma, chairman of India's nuclear regulatory board, says that the tsunami has not exposed any particular safety risks at Kalpakkam. He predicts that any necessary design changes could be implemented before the project's 2010 completion date.

Global geoscience suffers as UNESCO curtails funding

Quirin Schiermeier, Munich

Geoscientists are protesting against proposed cuts to a small but successful international Earth-sciences programme.

The International Geoscience Programme (IGCP), which provides seed money for local projects, has helped thousands of geologists from developing countries to coordinate their work and liaise with colleagues around the world. It has long been seen as the star of the Earth-sciences division of the United Nations Educational, Social and Cultural Organization (UNESCO).

But changes are afoot. When UNESCO's Earth-sciences director retired in November he was not replaced. And according to information leaked to *Nature*, UNESCO is to cut annual funding to the IGCP, currently about US\$200,000, by almost half from 2006. The IGCP should still receive about \$90,000 a year from the International Union of Geological Sciences.

Walter Erdelen, UNESCO's assistant director-general for natural sciences, confirmed that there are plans to substantially reduce the programme, although he would not say by how much. He says UNESCO's science activities are focusing on water and ecology, following a reduction in the total budget for 2006.

This focus is fine, says Sospeter Muhongo, a geologist at the University of Dar es Salaam in Tanzania and newly elected chairman of the IGCP's scientific board. But he adds that geologists would have much to contribute to such work.

Members of the IGCP's scientific board issued a joint communication last week appealing to UNESCO to maintain current levels of funding. "Drastic cuts would be demoralizing," says Muhongo. A final decision will be made at UNESCO's executive board meeting in April.

Since 1972, some 500 regional geological and mining-related projects in 150 countries have received seed funding of up to \$10,000 from the IGCP. Although other projects exist to help geoscientists in developing countries (see *Nature* 433, 449; 2005), geologists say that the community will be sad to see one of the most established programmes cut.

"Almost all of the impact will be on scientists from developing nations," says Douglas Erwin, a palaeontologist at the Smithsonian Institution's Museum of Natural History in Washington DC. "It is short-sighted for UNESCO to do this."

Additional reporting by Rex Dalton, San Diego.

Furore leads university to relent over Berlin science meetings

Munich The Free University of Berlin has backed down from its attempts to bring the prestigious Dahlem Conferences into more direct line with its own interests.

The conferences take place about three times a year in Berlin, and the proceedings are usually published rapidly. In the past year or so, the Free University, which administrates the meetings, has changed the organizational structure of the conferences, and dismissed the series editor (see *Nature* 433, 446; 2005). Protests at local and international levels have been vociferous.

Matters came to a head on 14 February, when more than half of the international scientific advisory board resigned. This followed the withdrawal of a conference from the schedule by its organizer, and confirmation from another organizer that the proceedings of a 2003 conference would be published elsewhere as the university had proved to be too slow.

The Free University now says that it will leave all scientific decisions to the scientific advisory board without undue influence, and will reinstate the dismissed series editor.

Ariane rocket launches its comeback campaign

Paris A weight was lifted from the shoulders of Europe's space community last weekend when its 'heavy lifter' Ariane 5 ECA rocket launched successfully.

This was the first time Europe's souped-up rocket had flown since a disastrous maiden flight in December 2002, when an ECA launcher veered off course and self-destructed, throwing the future of the rocket into doubt. Since then, the European Space Agency and its industrial partners have spent more than half a billion dollars on redesigning the launcher.

Taking off from the Kourou spaceport in French Guiana on 12 February, the Ariane 5



Back in business: the Ariane 5 ECA blasts off with its 8 tonne payload in French Guiana.

India offers Europe a ride to the Moon

Paris The European Space Agency (ESA) is considering an offer from India to place scientific instruments on its 2007 lunar mission, Chandrayaan-1.

ESA's Science Programme Committee, which discussed the plan last week, may spend up to €6 million (US\$7.8 million) on a package of instruments. European scientists would like to exploit the opportunity to analyse the chemistry of the Moon's surface using X-rays, to search for ice with infrared radiation, and to use visible laser light to create a topographical map. If the collaboration goes ahead, it would be the first time that Europe has joined a mission launched by the Indian Space Research Organisation.

ESA has also extended its current lunar mission, SMART-1, for a further year from August. SMART-1 was designed primarily to test new propulsion technologies but it also carried a



small scientific payload, including instruments that in January beamed back their first close-range images of the Moon's surface (above).

ECA put three satellites — a total of eight tonnes of payload — into orbit.

The rocket can actually lift ten tonnes, four more than a standard Ariane 5, and is intended to become the workhorse for future heavy payloads, including the James Webb Space Telescope and other deep-space probes.

Mystery deaths prompt Canadian ban for drug

Washington A drug used to treat attention deficit and hyperactivity disorder has been banned by Canadian authorities after only a year on the market.

Adderall, made by Shire Pharmaceuticals, based in Basingstoke, UK, was withdrawn after it was linked to 12 strokes and 20 unexplained deaths. Fourteen of the deaths and two of the strokes occurred in children. The US Food and Drug Administration issued an alert to health officials about the Canadian action but has not banned the drug. Adderall is not available in Europe.

Robert Peterson of Health Canada says that it cannot justify the risk of "very rare but nevertheless catastrophic adverse events".

Shire maintains that Health Canada's move is wrong. "We disagree with their action and we disagree with how they are interpreting the data. The general population rate of sudden death is greater than the rate of sudden death among the population taking Adderall," says spokesman Matthew Cabrey.

Capital proposal moves medical lab to city centre

London Britain's prestigious National Institute for Medical Research looks set to relocate from Mill Hill in the suburbs of London to a site close to University College

London (UCL) in the centre of the capital.

The Medical Research Council (MRC), the funding body that runs the institute, opted for the UCL site on 10 February after considering a rival bid from King's College London. The MRC says its decision hinged on UCL's expertise in taking basic research into the practical arena, and on the possibility of collaborations with physical and engineering sciences. A detailed proposal for the move will be considered by the MRC in May.

The move follows an acrimonious consultation process that set many institute staff at odds with senior MRC officials and that was last week criticized by politicians (see *Nature* 433, 564; 2005). The MRC says it will work closely with Mill Hill staff while the UCL bid is being developed.

Tech centre focuses on developing world

London The United Nations has unveiled plans for a major centre dedicated to studies of technology in poor nations.

The institute, which would bring together 100 researchers, will focus on economic and political studies of how new technologies can benefit less-developed nations. Two existing centres — the Maastricht Economic Research Institute on Innovation and Technology and the United Nations University Institute for New Technologies, also in Maastricht — will be merged this year to create the new body.

The two organizations have extensive experience of issues in the developing world, including work on building scientific capacity in Africa, studies of innovation in traditional agricultural practices and the development of policies to ally scientific research with local decision-making processes.

Setting sail for history

A small budget and big dreams make for a heady mix. But solar-sail pioneer Lou Friedman is ready for anything as spacecraft Cosmos 1 prepares to take on the Sun and the space agencies.

Tony Reichhardt reports.

This is a story about patience. Not just one man's patience, although Lou Friedman has waited half his life to get a solar sail into space. Futurists, too, have been dreaming about this technology for nearly a century and have yet to see it demonstrated. In April, if all goes to plan, a 600-square-metre Mylar sail called Cosmos 1, which looks more like a windmill than a starship, will prove that a spacecraft can be propelled by sunlight alone.

First, though, it will have to be launched into orbit on a converted missile from a Russian nuclear submarine in the Barents Sea. Cosmos 1 is privately funded by the Planetary Society, a US space-advocacy group based in Pasadena, California, which Friedman heads, but it was built in Moscow by the ex-Soviet aerospace company NPO Lavochkin.

After the sail reaches its initial 800-kilometre orbit and unfolds its eight triangular vanes, ground controllers will tilt the vanes like sailors feeling for the wind. A slight boost to the spacecraft's orbit is all they need to demonstrate propulsion by light pressure. It may take a few days, but the Cosmos team won't mind waiting.

Solar sails are not for people in a hurry. They accelerate almost imperceptibly at first, as photons of light bounce off their enormous reflective surface, imparting momentum. But, unlike conventional rockets, they can accelerate continuously, and keep accelerating as long as the Sun is shining, without needing a drop of fuel.

Light speed

After one day, the velocity increase for an interplanetary sail would be a modest 160 kilometres per hour. After 100 days, the sail would be moving at $16,000 \text{ km h}^{-1}$. In three years it would be travelling $160,000 \text{ km h}^{-1}$, three times faster than the Voyager spacecraft now exiting the Solar System, and fast enough to reach Pluto in less than five years — half the time the NASA New Horizons mission will take to reach Pluto.

This is why science fiction writers love solar sails — as do aerospace engineers, at least in theory. In the 1970s, Friedman was a project manager at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California, where he led the conceptual design of a US mission to Halley's Comet, using a gigantic $640,000\text{-m}^2$ sail. The idea was shot down by NASA management as too risky. "In retrospect it was too audacious," Friedman admits today, "and the schedule was unrealistic."

After leaving JPL, Friedman co-founded the Planetary Society with scientists Carl Sagan and Bruce Murray in 1980. While pushing the society's agenda of international space cooperation, he built solid friendships and working relationships with Russian space scientists and engineers, at a time when such relations were viewed with suspicion. Later, it made economic and technical sense for the Planetary Society to turn to Russia for help on the solar-sail project. The missile launch was a bargain; NPO Lavochkin was already working on inflatable spacecraft (the masts that hold the Cosmos 1 sail must inflate in space), and Russian interest in solar sails dates back to the visionary Konstantin Tsiolkovsky, who wrote about them as early as 1921.

Right now, Friedman is trying to balance his excitement at getting this far, with more realistic expectations. The project's tiny US\$4-million budget ("NASA would spend almost that much on the paper studies," he says) has meant cutting corners on certain materials and the number of design reviews. But the



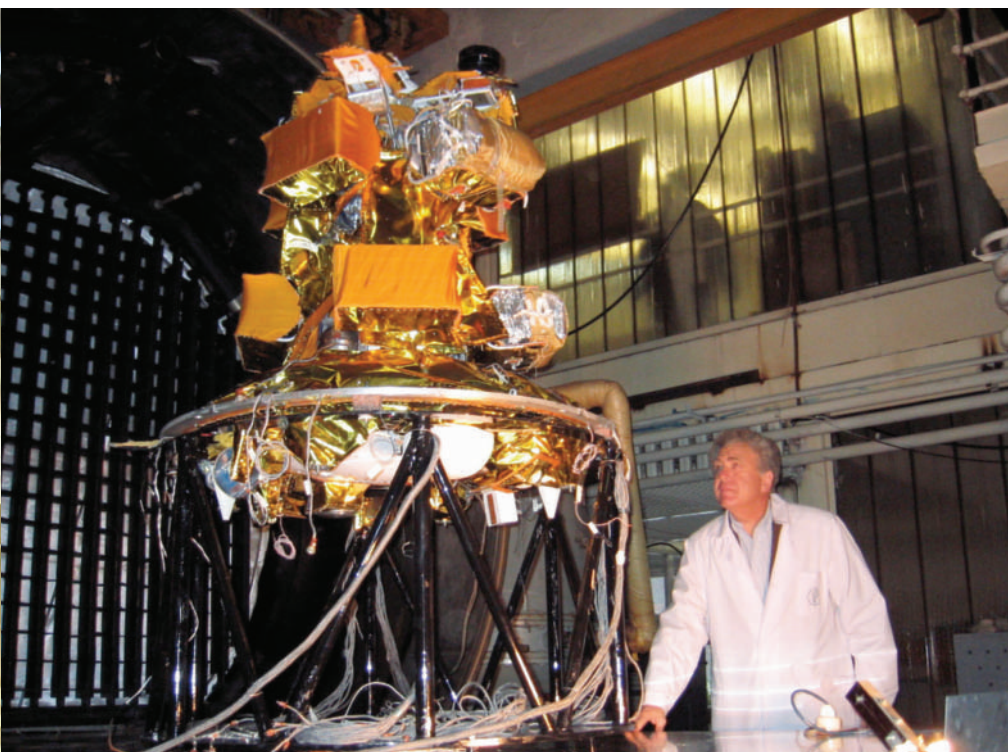
Engineers in Russia test the solar sails that will power a spacecraft.

team is made of "capable, experienced people" who have tried to anticipate everything that could go wrong. After one technical review in Moscow, recalls Friedman, a Russian consultant gave Cosmos 1 as much as a 70% chance of succeeding. "I said: 'You give it 70? Man, I'll take that!'"

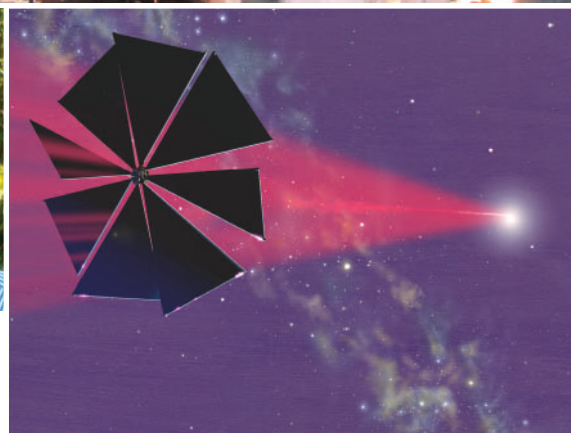
Unfortunately, the team missed its chance to test the sail's deployment on a suborbital launch in 2001, when the spacecraft failed to separate from its rocket, and both were lost at sea. Instead of repeating this short flight, Friedman and his colleagues decided to go straight to orbit for their next try.

Although NASA wouldn't approach a high-risk project in this way, the agency will be cheering Cosmos from the sidelines, says Tim Van Sant of the Goddard Space Flight Center in Maryland. Van Sant manages technology development for NASA's Sun-Earth Connection

"No one knows exactly how stable it will be, or whether it will twist and curl on itself like a flimsy kite in a strong wind."



Fact or fiction? Gregory Benford (above) plans to boost power to the solar sails designed by Lou Friedman (top) using lasers (right).



programme. His office has long contemplated solar sails for missions that cannot be done with conventional propulsion. Placing a spacecraft in a close polar orbit around the Sun, for example, requires enormous amounts of rocket fuel to fight the inward pull of solar gravity. Solar sails would act as natural brakes, and would never run out of fuel.

Catching the breeze

But there is a big gap between dreaming and doing. Most plans for solar sails have never got anywhere near the Sun. In 1992 there was talk of an international solar-sailing regatta to celebrate the 500th anniversary of Columbus's voyage to America. Groups including the Pasadena-based World Space Foundation got as far as building and testing sails on the ground, using a mix of professional and amateur labour, but their money dried up. More recently, a Texas-based group called Team Encounter announced plans to attach paying customers' messages, drawings, photographs and DNA samples to a sail and send

it off into interstellar space. None of these projects has come close to launching.

In terms of flight experience with solar sails, Japan is the world leader until Cosmos 1 launches. Last August the Japanese Aerospace Exploration Agency (JAXA) conducted a brief suborbital test on a sounding rocket, during which two 10-m sails unfurled from a mast to form a pinwheel shape. In May the agency will test a 20-m sail suspended from a scientific balloon at an altitude of 35 km.

JAXA's ultimate goal is a hybrid propulsion system combining solar sailing with ion-drive engines. One proposed JAXA mission would combine a 50-m solar sail with an ion drive to place a probe in orbit around Jupiter's poles and fly past several asteroids.

No one, though, has got as close to orbit as Friedman is today. Even if the sail deploys on cue, he's not sure when he should celebrate. "It could be a very tenuous success," he says, as the mission team tries to control its sail and push it into a higher orbit. Of the many technical challenges facing solar sails, Fried-

man counts dynamics as perhaps the most vexing. No one knows exactly how stable it will be, or whether it will twist and curl on itself like a flimsy kite in a strong wind.

If Cosmos 1 reaches a higher orbit and meets all its other mission objectives, it will be used for one last, futuristic experiment. University of California physicist and science fiction writer Gregory Benford, along with his brother James, president of Microwave Sciences in Lafayette, California, will aim a 450-kilowatt microwave beam from a radio antenna in the Mojave Desert towards the sail. They hope the beam will give an extra push to the sail. Someday, that method may be used to propel gossamer sails more quickly to other planets, and perhaps even other stars.

On the right tack

A successful Cosmos 1 mission would give a gentle push to solar-sail projects within the space agencies. The earliest NASA could fly a solar sail is 2009 on the Space Technology 9 demonstration mission, although other technologies will be competing for that flight.

Among NASA's long-term solar-sail proposals are the Particle Acceleration Solar Orbiter, which would orbit close enough to the Sun to keep a steady gaze on active solar regions, and a Solar Polar Imager for studying the Sun's higher latitudes. NASA and the National Oceanic and Atmospheric Administration would also like to put weather stations in stable orbits between Earth and the Sun to give advance warning of sunstorms. These would need sails three to five times bigger than Cosmos 1.

The European Space Agency is interested in solar sails too, for reasons similar to NASA's — to place an orbiter around the poles of the Sun. The agency is also studying another mission concept called Earthguard to visit near-Earth asteroids.

Although no solar-sail missions have yet been approved, Van Sant is spending about \$10 million a year laying the technological groundwork. This year two pioneers of solar-sail development, L'Garde of Tustin, California, and ABLE Engineering of Goleta, California, will test different designs for 20-m sails in a giant vacuum chamber at NASA's Plum Brook facility in Ohio.

ABLE's sail is made of an extremely thin new polymer called CP-1, only 2.5 micrometres thick — half as thick as the aluminized Mylar on Cosmos 1. "If you sneeze, you'll send this stuff across the table," says Van Sant. Thinner and lighter is better when it comes to solar sails, but flimsy films are also more prone to tearing. Even the reinforced sails of Cosmos 1 won't last forever: within a month of launch they will begin to degrade in the harsh sunlight. But the short flight should be long enough to demonstrate the principle of solar sailing, and if successful will open the heavens to other solar-powered craft. ■

Tony Reichhardt writes for *Nature* from Washington DC.

Hopping fences

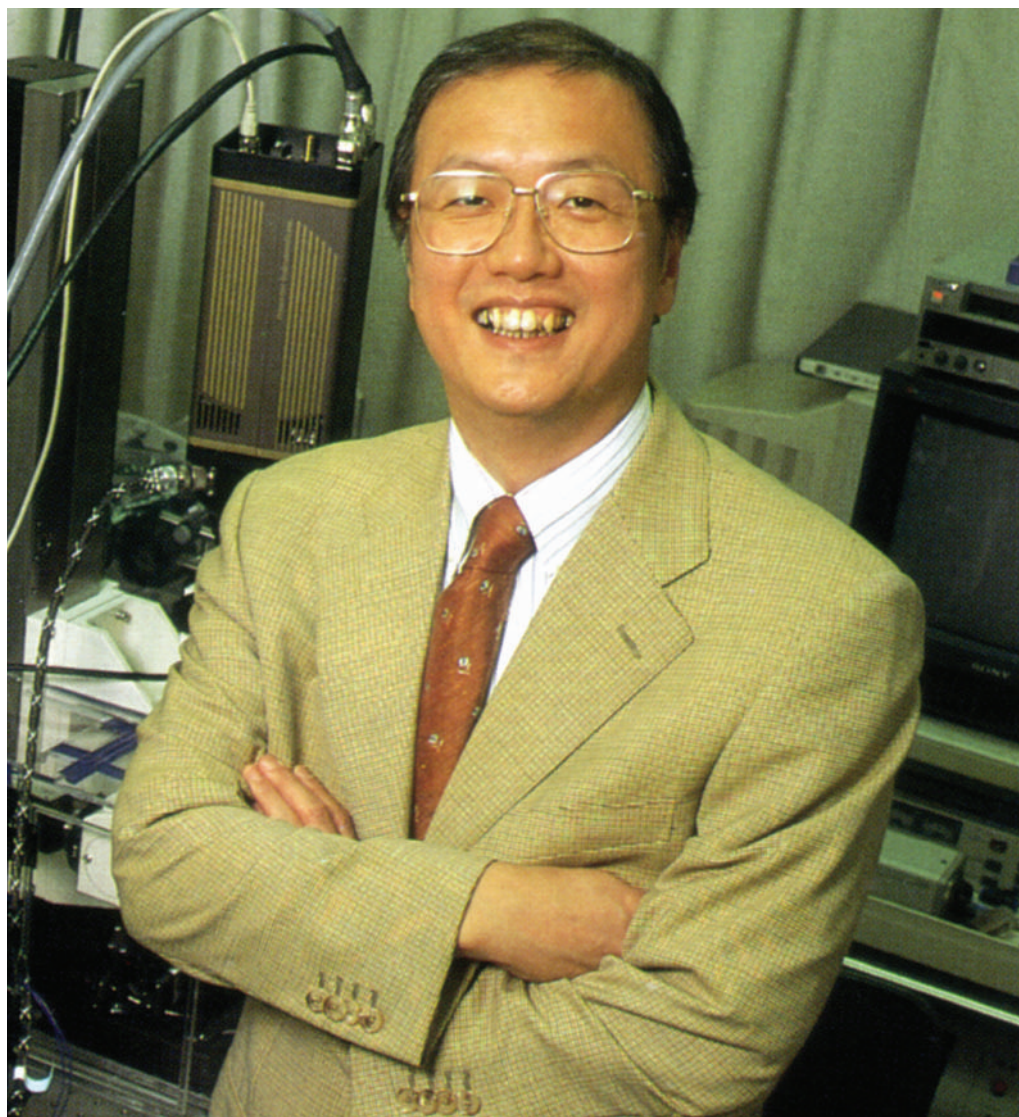
Just how proteins jostle around in the fatty membrane that surrounds every living cell has been a source of debate for decades. Now one researcher is using an ultra-high-speed camera to watch this dance in unprecedented detail — but that hasn't stopped the arguments. Alison Abbott investigates.

When Akihiro Kusumi ran his grainy black-and-white film of a single molecule dancing maniacally against a grey background, a wave of confusion passed through the audience. Those seeing the pictures for the first time gaped in amazement and scepticism. Questions were fired from the floor. What exactly is being measured? How did you do it? Can you really catch the rapid movements of such a molecule on film?

At this conference in Italy last autumn, Kusumi was claiming to show the movement of a protein within the lipid membrane of a living mammalian cell, on microsecond and nanometre scales. In decades of membrane research, no one else has come close to such a precise level of visualization.

Kusumi's analyses have led him to propose a sweeping hypothesis for cell membrane dynamics, which he calls 'hop diffusion'. He suggests that molecules move around in the membrane at breakneck speed, but are confined in small areas by 'fences' created by the cell's actin cytoskeleton — a mesh of filaments that helps give a cell its shape. Occasionally the molecules hop over these fences, and so seem to travel over larger distances at a relatively sedate rate (see 'Fenced in', opposite).

This rather simple idea may not sound



Caught on camera: Akihiro Kusumi believes he has filmed proteins moving in cell membranes.

like a bombshell, but it is causing a sensation in the world of membrane biophysics. If true, the concept would solve some long-standing puzzles about how proteins move within a membrane. Perhaps more important, it would also shed light on how signals from the outside world are transmitted through the membrane to the inside of the cell — a process that is central to cell behaviour in health and disease. Working out the details of signal transduction is one of today's major challenges. Cell and molecular biologists have spent years identifying which proteins in a cell membrane need to bind together to convey a signal, but they have not been able to explain how these proteins manage to meet and interact so successfully.

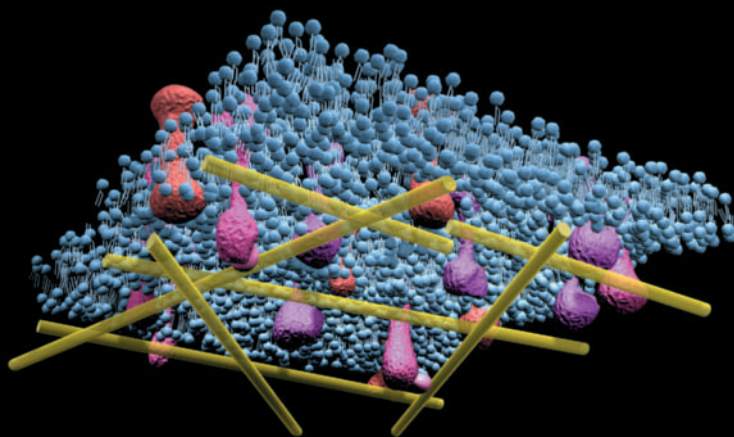
Kusumi's work has been controversial among the *cognoscenti* for some years, and in the past he has had problems getting published. This was partly a consequence of his interdisciplinary approach, which brought into the field new technologies that no one else was able to acquire — most pertinently ultrafast filming technology borrowed from

the field of explosion research. In the past few years his work has gained greater acceptance, although there are still those who argue with his bold interpretations of his data.

Movie maker

Kusumi, who has recently moved to the University of Kyoto, is relentlessly cheerful, and seems to take attacks on the chin, ploughing happily on and finding humour in every situation. Like the molecules he studies, he does not like to be too confined. And he certainly does not want to be placed in a tight scientific box. "I studied both physics and biology as an undergraduate and did my PhD in biophysics," he says. "People like to think that I spend most of my time working with machines, but in fact I spend most of my time thinking about biology." Such credentials, he indicates, have helped him to bridge the parallel worlds of biophysics and cell biology.

He began his independent research career in the early 1980s. Then, the working hypothesis for membrane dynamics was the



Cell membrane structure: a variety of proteins are held in a double layer of lipids (blue). It is possible that the underlying actin cytoskeleton of the cell (yellow) confines the movement of the proteins.

1972 'fluid mosaic model', which described membranes as a double layer of lipids that holds protein molecules (see picture, above). Some of the proteins pass from outside the cell right the way through the 'lipid bilayer' into the cell's inside; others sit in just the outer lipid layer poking their heads outside the cell membrane. Yet others nuzzle onto or into the inner layer of lipids from within the cell. The membrane proteins are not fixed but, according to this theory, can diffuse freely around in their lipid environment through brownian motion — the random movement of atoms and molecules due to thermal energy.

This structural picture, which is still believed to be broadly true, is supported by biochemical and microscopic evidence¹. But there are serious problems with the idea that brownian motion alone is responsible for the movement in a cell membrane.

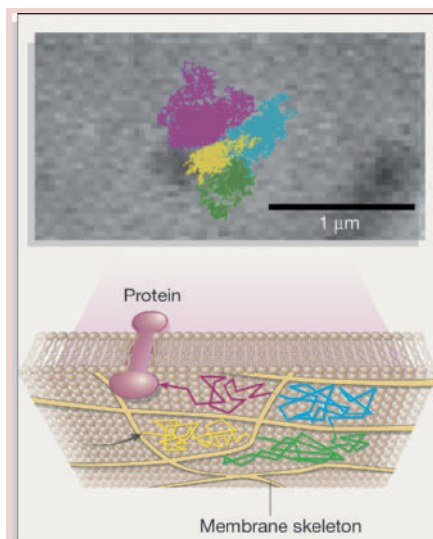
In the 1980s, it became clear to biophysicists that proteins in living cells moved within the membrane up to 100 times slower than they did in artificial membranes — simple lipid constructions made in the lab. Something, they reasoned, must be constraining the free diffusion of proteins in natural, living membranes.

Hidden depths

Cell biologists, meanwhile, had a problem with the very concept of proteins moving by free diffusion — it didn't square with new information about the function of membrane proteins. Many membrane proteins are involved in the complex business of cellular signalling. A common signal is the binding of a hormone or a growth factor to the part of a membrane-spanning receptor protein sticking outside of a cell. The binding causes the receptor to change shape, sparking a sequence of interactions with other proteins in the lipids that allows

the message to be carried to the appropriate machinery inside the cell.

Cell and molecular biologists identified a large number of such proteins in different signalling pathways, and by the 1990s they were no longer convinced that free diffusion was sufficient to explain how the right line-up of proteins arrived in the right place at the right time to guarantee interaction. It was becoming clear that a higher order of organi-



Fenced in

Akihiro Kusumi's hypothesis proposes that the cell's internal skeleton (yellow) acts like a fence on the inner side of the cell membrane. A protein races around within one fenced-off section at high speed, before hopping the fence into another section. The evidence is shown in the photo above, which combines tens of thousands of snapshots to show the protein's location over a period of a second or so. Kusumi has joined the dots of the protein's journey with a colour for each fence-posted area.

zation in the membrane was called for to constrain the signalling molecules.

Enter the 'lipid raft hypothesis', which proposed that protein movement could be constrained by the physicochemical properties of different mixtures of lipids. According to this idea, the various lipids in a cell membrane are not randomly distributed, as previously supposed, but can group into clusters that are more viscous than the rest of the membrane². These floating rafts of specific types of lipid provide favourable environments for certain proteins, and so effectively confine them to a limited area. The coming together of several small rafts during signalling initiation would then get all the right proteins into the same place at the same time.

The raft hypothesis was a godsend for those developing ideas about the dynamics of cell signalling. But there are sceptics. Although rafts have apparently been seen under the microscope, most evidence for their existence is indirect, and some believe that they are an artefact of the way the cells are prepared for analysis. Someone has even coined the term 'unidentified floating objects' for them.

Cells on film

But if rafts are not the answer, what else could confine protein molecules to a small space for long enough to allow the appropriate interactions to take place? Kusumi decided that the only way forward was to study the movement of single molecules. He found a way to attach a colloidal gold particle, 20–40 nanometres in diameter, to a membrane protein and video its movement. Using a rat kidney cell he made his first observation of apparent hop diffusion³. The gold particle whizzed around in a small area of the membrane and then seemed to jump into a neighbouring area, rather than moving randomly at a more sedate speed throughout the whole membrane or being confined to a raft.

He reasoned that the compartments could be defined by fences made from actin filaments of the cytoskeleton within the cell pushing against the underside of the membrane. This concept extended the theoretical work of Michael Sheetz, a biophysicist at Columbia University in New York, who in 1983 had first proposed that transmembrane proteins might be constrained by the cytoskeletal network in red blood cells⁴.

But when Kusumi looked for hopping in many other types of cells, including red blood cells, he couldn't find it. His data showed proteins that seemed to move freely around the membrane, without lingering in areas between supposed fences. He knew that his camera's spatial resolution was good enough to pick up the expected jittering movement within confined areas; but maybe, he thought, it wasn't quick enough.

To find out, he needed a faster camera —

much faster. "I was certain that explosion researchers would be using very fast cameras. I remembered that scientists developing the atomic bomb in the Manhattan project used them to observe explosions," he says. He found several modern models on the market, but they were for areas such as explosion research and were not specifically designed for microscopic work.

Kusumi acquired one anyway, which at 40,000 frames per second gave a 25-microsecond resolution — 1,000 times faster than his previous camera. "We had to integrate the camera with our microscopes and then optimize the system through a lot of tinkering," he says. 'Tinkering' is too small a word for what was required, which included improving contrast, introducing a light source that kept the camera, the microscope and the cells happy, and, most important, speeding up data analysis. Kusumi says that without his multidisciplinary team of 20 researchers, ranging from computer scientists to physicists and biochemists, it would not have been possible to create the complex machine that takes his apparently simple films.

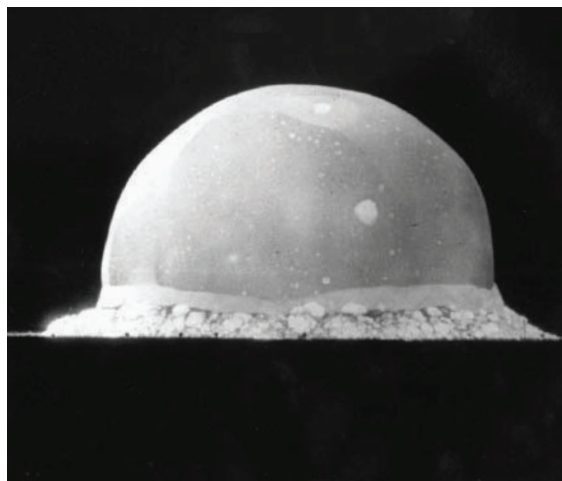
Hop and glory

With the new system in place, his team was soon seeing what looked like hop diffusion in many different types of cell. On average the group would see a hop every 1–1,000 milliseconds and membrane compartments of about 30–200 nanometres across. This explained the difference between diffusion speeds in natural membranes and artificial ones. Proteins move at the same fast speed in both, but their progress is hindered by fences in the natural system, making their long-distance diffusion slower. "The hops were being missed in video recording, which has a resolution of only 30 milliseconds," says Kusumi. "A lot can happen in 30 milliseconds."

"We were delighted, but we still knew that an enigma remained," says Kusumi. It is not just proteins within the membrane that move — the lipids move too. And these also seem to diffuse more slowly in natural cells than in artificial systems, even though the outer layer of lipids is well out of reach of the cell skeleton underneath.

It proved much harder to label and track lipids compared with proteins. But after two years and thousands of films, Kusumi succeeded in getting statistically sound results, showing lipids hopping too⁵ — they move, he found, at a rate of one hop every 1–100 milliseconds. This, he thinks, is because some proteins that pass right through the membrane are anchored against the cytoskeletal fence, and so act as fence posts or pickets. These mark out corrals that help to confine the membrane lipids.

In 2001, Kusumi began to present his results, his interpretation and his general view of how a membrane keeps its house in order. At the same time, other work from his lab was adding weight to his theory⁶. Experiments with electron microscope computed tomography of the cytoskeleton, sophisticated spectrometers and other high-tech instruments such as molecular tweezers⁷, he says, corroborated the size of the fenced-in corrals. And, Kusumi adds, the force required to drag a molecule over a fence is in line with what the hypothesis would predict. What he did not see was



A nuclear explosion after 0.16 seconds, caught on camera in 1945. Similar techniques are now being used to image cell molecules.

evidence that rafts played any part in initiating cell signalling.

Kusumi's grand hypothesis, created with self-assembled high-tech equipment, really shook things up. "There is no doubt that Kusumi has had a dramatic effect on the field of membrane biophysics," says Ken Jacobson, a cell biologist at the University of North Carolina at Chapel Hill. "He has pushed the field on several fronts."

Jacobson, who also tracks single particles in cell membranes, but using a slower-speed video camera, comments that Kusumi's hop-diffusion model has been accepted to a "modest degree" in the field. One of the major obstacles to full acceptance, he says, is that no other lab currently has the high-speed equipment to confirm the results independently.

Golden retriever

Another worry is the use of a relatively large gold particle for tracking — a particle that may be larger than some of the corrals. Kusumi argues that he carries out extensive, time-consuming control studies to ensure that the dot of gold is pulling only a single molecule. In his research community Kusumi is known for his obsessive rigour — most people have the greatest confidence in his technical prowess, says Jacobson. "But even so, you always have that nagging doubt

that somehow a 40-nanometre particle may be affecting the results in some way."

There are rivals to the fence-and-picket model that might explain the curious dynamics that Kusumi observes, adds Jitu Mayor, a cell biologist at the National Centre for Biological Sciences in Bangalore. Some scientists argue that the simple crowding of proteins in living membranes could account for the relatively slow protein movement, he says. But Mayor argues that no one else has tackled the problems as thoroughly as Kusumi, coming up with a broad, internally consistent hypothesis and then working to validate it. "The rivals just don't make testable predictions," he says.

Kusumi is not only dedicated to putting his theory through the wringer of experiment; he is also willing to sit on the fence over the question of rafts. Once a cell has begun the process of cell signalling — after the proteins required have already come together — he says he does see some evidence for the raft hypothesis. "We see molecules forming clusters when they are stimulated by a signal, so they can be seen as rafts," he says, referring to unpublished work. "But we don't see them in non-stimulated cell membranes."

Kai Simons, a director at the Max Planck Society for cell biology in Dresden, Germany, and one of the strongest proponents of raft theory, is mollified by this. If Kusumi can't see rafts before cell signalling, says Simons, perhaps that's because the gold particle is perturbing the system too much. In the case of stimulated cell membranes, he says, he and Kusumi are happily in agreement.

Kusumi takes his critics seriously. But he laughs about it too, throwing out asides: "new ideas can be difficult," he says, and "physicists are fond of brownian motion."

So Kusumi will forge ahead, extending his work to look at the interactions between cell membrane proteins and nanomaterials being developed for tissue transplantation. He has just moved from the University of Nagoya to a new centre, whose name reflects his tendency to work at the edge: the University of Kyoto's Institute of Frontier Medical Sciences. It isn't the first time Kusumi has shifted his workplace — he has moved once every eight years to a new city. "I lead my life according to hop-diffusion principles," he laughs. Thankfully, he's easier to catch on camera.

Alison Abbott is Nature's senior European correspondent.

1. Edidin, M. *Nature Rev. Mol. Cell Biol.* **4**, 414–418 (2003).
2. Simons, K. & Ikonen, E. *Nature* **387**, 569–572 (1997).
3. Kusumi, A., Sako, Y. & Yamamoto, M. *Biophys. J.* **65**, 2021–2040 (1993).
4. Sheetz, M. P. *Semin. Hematol.* **20**, 175–188 (1983).
5. Murase, K. *et al. Biophys. J.* **86**, 4075–4093 (2004).
6. Kawasaki, K., Yin, J. J., Subczynski, W. K., Hyde, J. S. & Kusumi, A. *Biophys. J.* **80**, 738–748 (2001).
7. Nakada, C. *et al. Nature Cell Biol.* **5**, 626–632 (2003).

No political will to seek innovative contraception

Focus instead on other reproductive issues, which may make birth-control superfluous.

Sir—Your Outlook supplement on fertility includes a Commentary article by J. F. Strauss and M. Kafrissen, “Waiting for the second coming” (*Nature* **432**, 43–45; 2004), which is either an extraordinarily naive recommendation for developing fundamentally new “safer, more effective and more user-friendly” contraceptives or a disingenuous plea for more money. In this field, pouring additional funds into research without addressing the underlying problems is as illogical as getting nine women pregnant and expecting a baby in one month.

The opening sentence, “Even developed countries have a staggeringly high incidence of unplanned pregnancies”, leads the authors to conclude that currently available contraceptives “are simply not meeting the needs of society”. Yet it is primarily through the use of the Pill, intrauterine device (IUD), condom, sterilization and abortion that the reproduction rate in every European country except Albania and Malta is now below replacement level. Even in Japan, where the Pill was legalized only in 1999, the rate fell below replacement level years ago. A combination of almost static contraceptive ‘hardware’ and the more dynamically changing ‘software’ — the cultural, economic, public health, educational, political (notably in

China) and women’s-rights issues — has proved crucial to the decline in much of the world’s population growth.

Hence it is shameless grantsmanship to claim that promising leads will result from more money for “forays into ‘glycomics’ and ‘lipidomics’ [and] new technologies for large-scale analysis of carbohydrates and lipids”. Such leads will have nothing to do with practical contraception within the reproductive lifetime of any living infant.

Of the 20 largest pharmaceutical companies in the world, only two (Johnson & Johnson and Wyeth) market female contraceptives and do a modest amount of R&D to improve existing steroid contraceptives. None of the 20 is active in the male-contraceptive field. None has been willing to continue contraceptive-vaccine research initiated by the World Health Organization. The types of incentives listed by the Commentary authors to encourage a re-entry of the pharmaceutical industry, now fixated on blockbuster drugs for geriatric populations, are largely a rehash of earlier proposals (see C. Djerassi *Science* **245**, 356–361; 1989) for which there is no political will for implementation.

There are no short cuts to establishing the authors’ illusory goal of “safer” birth

control (safer than condoms, IUDs or sterilization?) without demonstrating clinically that consumption of a fundamentally new contraceptive for 10–20 years by women or 20–40 years by men is indeed safe. Who will pay the hundreds of millions of dollars required for such studies?

It makes more sense to focus on the reproductive realities described by other articles in your Fertility Outlook: research on conception, infection and the extension of female fertility. Or the enormous practical advances made in assisted reproduction, which do not require the pharmaceutical industry for research or practical implementation. Consider that in China and the United States, sterilization has now surpassed the Pill and other contraceptives among married couples.

As sex and fertilization become increasingly separated, the cryopreservation of semen by young men followed by sterilization and subsequent use of artificial insemination to create their future one or two children (see C. Djerassi and S. P. Leibo *Nature* **370**, 11–12; 1994) may one day make contraception superfluous.

Carl Djerassi

Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA

Emissions control needs atmospheric verification

Sir—As your News Feature “The carbon game” (*Nature* **432**, 268–270; 2004) makes clear, the start of the European Emissions Trading Scheme may be a route to controlling greenhouse gases. Coupled with the United Nations’ Clean Development Mechanism to encourage participation by the poorer nations, it may indeed work.

However, independent atmospheric verification of emissions has long been neglected. The system relies on producers reporting their own greenhouse emissions. Until now, there has been no financial penalty for producing emissions and no benefit from carbon sequestration. Now that money enters the picture, so also can fraud.

Today’s emitters of greenhouse gases report their emissions; the information is then passed on to national and international databases. Under the new scheme, there will now be an incentive to under-report emissions at every stage. Equally, those sequestering carbon will be tempted to exaggerate. Despite diligent

independent controls, doubts will arise at all levels. Even when there is no intent to misreport, companies, regions and nations will suspect each other.

“Trust but verify” as Ronald Reagan used to say, quoting a Russian motto. The Nuclear Test Ban Treaty created trust by diligent verification. Atmospheric monitoring detected fallout, and a global network of seismographs was set up in order to detect explosions. Trust came, bomb tests ended. Plate tectonics was a surprise bonus.

The Kyoto Protocol lacks this. Verification is not expensive, but atmospheric monitoring at present is inadequate. A modest but effective multinational programme to assess net carbon-gas emissions would not cost much. Satellites give broad-brush imagery, but the main need is for very precise measurements on the ground, which are not expensive to carry out. For methane, modelling based on careful monitoring of concentrations and isotopes is already verifying national emission inventories by source type and seasonality.

Most of the monitoring load falls on the United States. There are a few

exceptions in the form of some excellent national programmes, and the European Union has programmes such as Carbo-Europe. But many non-US programmes have only short-term funding, despite the need for continuity. The British contribution is minimal. There is little on the tropical landmasses.

Other nations must do more to share in detailed long-term understanding of carbon budgets, including uptakes by the terrestrial biosphere. The Commonwealth ‘club’, led by strong Australian and Canadian programmes, could help in the tropics.

To win the trust of the United States, China and India, whose emissions growth will damage us all, we need to verify emissions and uptakes accurately and in detail.

Euan Nisbet

Atmospheric Laboratory, Department of Geology, Royal Holloway, Egham, Surrey TW20 0EX, UK

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

In search of weighty matters

A hard look at the 45-year quest to detect gravitational radiation.

Gravity's Shadow: The Search for Gravitational Waves

by Harry Collins

University of Chicago Press: 2004. 864 pp.

\$100, £70 (hbk); \$39, £27.50 (pbk)

Virginia Trimble

A distinguished colleague advised me not to review Harry Collins' book to "avoid the controversy". I think, however, that no discernible controversy remains. The physics and astronomy community has recognized for 30 years that gravitational radiation exists but that Joe Weber didn't discover it (roughly the first half of the book). And in the past decade, a particular sort of detector, called a free-mass interferometer, has pretty much wiped the competition off the face of the Earth (the second half of the book). The best-known and most expensive (well, it's American) version of the winner is called LIGO, the Laser Interferometer Gravitational Wave Observatory. Collins describes carefully how it won out over bar and sphere detectors and how it has (so far) been designed, managed, built and debugged. The single most salutary lesson he draws is that "the more money you want to spend on science, the less expert will the decision makers be".

What is gravitational radiation? Well, to produce light (electromagnetic radiation) you wiggle electric charges around; to produce gravitational radiation you wiggle masses around. There are technical differences in the polarization, for example, but the main difference is that gravitational radiation is the weaker by a factor of 10^{40} . This means the masses that wiggle have to be of astronomical size to stand any chance of detection, so billion-dollar projects are currently under way to try and achieve this.

Collins is director of the Centre for the Study of Knowledge, Expertise and Science at Cardiff University, UK. The search for gravitational radiation (waves in current parlance; he explains why) has been his most abiding interest, with the interviews reported in the present volume dating from 1972 to 2003. The book is not very easy going for a physical scientist. I caught at least 50 technical terms and concepts whose meanings I could not have guessed or would have guessed wrong, because (like energy and momentum in physics) they are everyday words used in non-everyday senses, such as moral integration and colonial cringe.

Is it worthwhile for a natural scientist to master some of these? Probably. Ornithology, *pace* Richard Feynman, is useful to birds through the designation and protection of



Looking for clues: Joe Weber used bar detectors in the 1960s to search for gravitational radiation.

endangered species, although the birds don't know it.

Collins coins the term 'Pascalian funding', for instance, to describe why government agencies in 1984 began supporting Weber's work on coherent neutrino scattering (with principles later applied to bar detectors for gravitational waves). The idea is that some things would be so important if they worked that investment makes sense even if there is only a very small probability of success. Pascal had in mind belief in a deity. The coherent neutrino process would have made submarines clearly visible under the oceans via emission from their nuclear reactors, though the explanation given of the physics is unclear because the author fails to mention the closest working analogy, Mossbauer scattering of gamma rays. The Star Wars missile defence system and cold fusion, whose funding has puzzled many physicists, probably also come under the Pascalian rubric.

Collins' discussion of evidential collectivism and individualism should ring bells with astronomers who have noticed the very different data release patterns of the Sloan Digital Sky Survey (let it all hang out) and the Wilkinson Microwave Anisotropy Project (release nothing until the team agrees, and then only a final product) and supposed that one must surely be wrong. Both can be right. The Sloan team are evidential collectivists who wish to involve the wider community in the process of assessment from an early stage, while the Wilkinson team are evidential individualists who believe they should take full

responsibility for the validity and meaning of a result before it leaves their laboratory.

New in this book is a metaphor of a dam of generally accepted ideas and experimental results holding back a lake of active work, within which a small island holds a few unconventional workers, Weber of course among them. Add the now common phrase "voted off the island", and you will have a fair picture of what happened. Between about 1972 and 1976, most of the physics community decided that the early reports of evidence for gravitational radiation could not have been right, and thereafter reacted to publications and conference talks by Weber and his colleagues as if they simply did not exist.

The author misses some points. A footnote mentions that the two messiest offices he ever saw were those of Weber and of Ron Drever, architect and major builder of the 40-metre prototype for LIGO, who was summarily removed from the team in 1992. Collins seems not to have realized that each had been suddenly evicted from a large laboratory and was required on almost no notice to squeeze many years of notes, records and spare parts into an average-size faculty office.

Contributory expertise is what is needed to do science. Interactional expertise is sufficient for communicating about it: "You listen a good low-temperature physics," as Lev Landau once said to the wife of a colleague. Collins describes himself as having interactional expertise in gravitational-wave detection, but might have benefited from additional advice by contributory experts.

My teeth were set on edge by reference to “the stable form of uranium”, a violation of Kepler’s second law in a description of how the Earth’s orbit would change under various circumstances, and by “the rest mass of the neutrino is 4 eV”.

Collins has been well served by his editor and publisher, but not perfectly. There are un-sort-out-able mismatches between text and index, references and figures; acronyms in the second half of the alphabet go undecoded; several well-known names are misspelled. And readers are informed that Weber’s death occurred “on September 31, 2000”. Well, Joe always said he could do things that other people couldn’t, but there are limits.

Incidentally, my adviser was partly right: I should not have agreed to review this book. It is very much harder to hear harsh, sometimes false, things said about one’s spouse after he can no longer defend himself. I am not alone in this feeling. Carvel Gold, widow of Thomas Gold, whose work was also far from universally accepted (see *Nature* 430, 415; 2004), says the same thing.

Virginia Trimble is at the University of California, Irvine, California 92697-4575, USA. She and Joe Weber were married from 16 March 1972 until his death on 30 September 2000.

Watson’s way with words

The Writing Life of James D. Watson

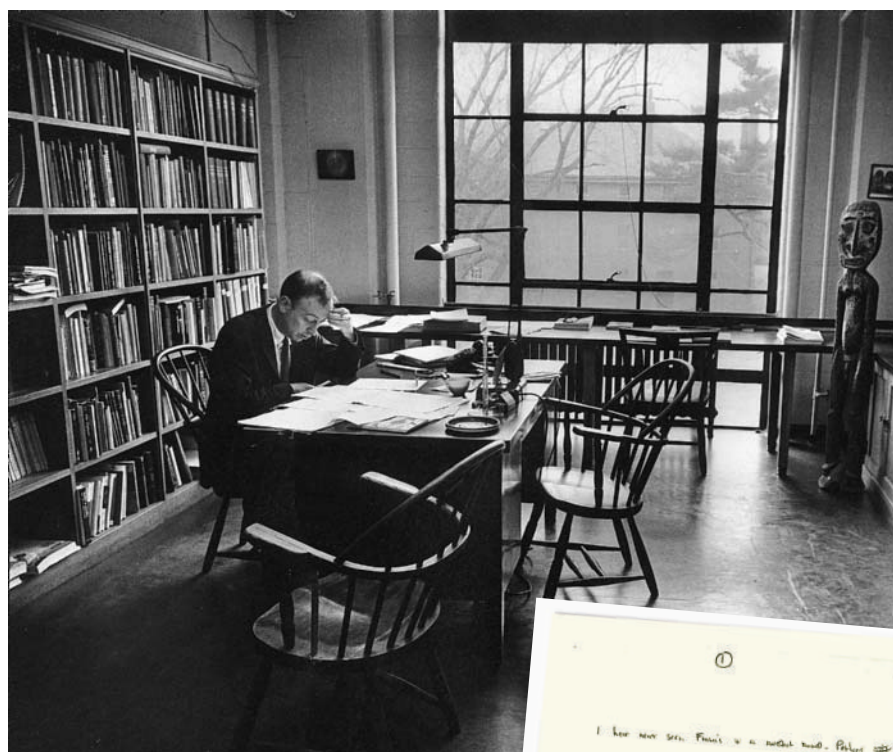
by Errol C. Friedberg
Cold Spring Harbor Laboratory Press: 2005.
193 pp. \$25, £18

Lewis Wolpert

The Double Helix would on its own have established James Watson’s reputation as a writer: it is the only book about science to appear in both the board’s and the readers’ lists of the Modern Library’s top 100 non-fiction works. But Watson’s textbooks have also given scientists, particularly students, a deeper understanding of genes and cells. And his popular-science books have given the public a new image of scientific research. *The Writing Life of James D. Watson* examines these achievements.

Watson was brought up to believe in the importance of books and reliable knowledge. He read widely and particularly enjoyed books by Graham Greene and *Arrowsmith* by Sinclair Lewis. Reading Erwin Schrödinger’s *What is Life?* at the age of 17, Watson became convinced that genes were the essence of life and decided devote his own to their study. By the age of 25 he had, with Francis Crick, discovered the double-helix structure of DNA.

In relating this story in *The Double Helix*,



The write idea? In *The Double Helix*, James Watson gave a personal account of the quest for the structure of DNA.

Watson set out to produce a good story that the public would enjoy as much as *The Great Gatsby*. He started writing in 1962 with the working title “Honest Jim”, which is illuminating in itself. *The Writing Life of James D. Watson* includes images of both the handwritten manuscript and the galley proofs. Indeed, almost half of Friedberg’s book is devoted to photographs of text and letters, and of Watson and friends — they take up too much space, I think.

A draft of *The Double Helix* sent to Crick and Maurice Wilkins, a co-discoverer of the double helix, began: “I have never seen Francis Crick in a modest mood.” This upset them so much that they threatened legal action. Harvard University Press was due to publish the book, but concerns about its libellous potential, and Watson’s refusal to change the text, caused them to withdraw, so Atheneum Press published it instead. Watson was delighted that Lawrence Bragg agreed to write a foreword.

The great X-ray crystallographer J. D. Bernal could not put the book down, but thought it was particularly unfair to Rosalind Franklin. Initial reviews were mixed, but Peter Medawar wrote that “it will be an enormous success, and deserves to be so — a classic in the sense that it will go on being read.” He was, as usual, right. Yet Crick found it difficult to take Watson’s account seriously, although he did appreciate the quality of the writing.

Watson’s skill as a writer is illustrated by this description of Rosalind Franklin. “Though her features were strong, she was

not unattractive and might even have been quite stunning had she taken even a mild interest in clothes. This she did not. There was never lipstick to contrast with her straight black hair, while at the age of thirty-one her dresses showed all the imagination of English bluestocking adolescents.” Would that other scientists could write as well as that. Of his later memoir *Genes, Girls and Gamow*, some said that his style broke new ground with its postmodern innovatory syntax, but others were critical of both its literary style and its content. It is a pity that there are few examples of Watson’s writing in Friedberg’s book, nor any real analysis of the way he writes.

In all Watson’s writing — from director’s reports for the laboratory at Cold Spring Harbor to the popular-science books *A Passion for DNA* and *The DNA Story* — his strong character emerges: his sarcasm, criticism and praise make it clear what he thinks. His love of science and DNA always comes through, as does his contempt for his enemies.

Science in culture

Eastern promise

The Queen Anne churches in east London were precisely aligned on an east–west axis.

Heike Langenberg

In 1714, Edmond Halley, professor of geometry at the University of Oxford, added the finishing touches to his latest paper, on “the variations of the magnetical compass” (*Phil. Trans. R. Soc. Lond.* **29**, 165–168; 1714). Halley, best known today for the comet that bears his name, was interested in a wide range of fields, including astronomy, meteorology and geomagnetism.

Around the same time, the foundations of two new churches were laid in east London: Christ Church Spitalfields and St Anne’s, Limehouse (right). Both buildings were aligned with remarkable precision on an east–west axis.

In a forthcoming paper in the *Journal of the Society of Architectural Historians* (**64**, 56–73; 2005), geologist Jason Ali and historian Peter Cunich suggest that Halley put his work on declination-corrected compasses into practice in surveying the two churches. They argue that this may have been the first time that this modern science-based technique, which corrects for the difference between magnetic and true north, was used to align buildings.

Christ Church Spitalfields and St Anne’s, Limehouse, were constructed under a programme established by two acts of parliament in 1711 and 1712 to build 50 churches in London’s rapidly expanding East End. The programme was an attempt to fight irreligion and the rise of the dissenting churches in the poor suburbs.

The parliamentary acts were passed in the reign of Queen Anne when the Tory party was in power. It was in the Tories’ interest to strengthen the Church of England because the parliamentary opposition party, the Whigs, were supported by religious dissenters. The buildings were specifically commissioned to be aligned in the traditional way along an east–west axis, emphasizing the link between the Anglican church and early Christianity.

In the end, only 12 churches were built, partly



owing to financial constraints and partly because the political balance changed after the death of Queen Anne in 1714, when the Whigs became the dominant party.

The commission for building the new churches appointed Nicholas Hawksmoor as one of its architects. Hawksmoor had been a student of Halley’s fellow Royal Society member Christopher Wren, who is most famous for designing St Paul’s Cathedral. Halley and Hawksmoor had shared a circle of friends since the 1680s.

Traditionally, the alignment of churches had

been achieved by using the position of the rising or setting sun on particular days: Easter, the feast day of the church’s patron saint, or one of the equinoxes. But this approach cannot explain the precision in the alignment of Christ Church Spitalfields and St Anne’s, Limehouse.

At London’s latitude, the seasonal range of sunrise is quite large, from 51.5° at the summer solstice to 128.5° in mid-winter. Even a few days before or after the equinoxes, the sun rises and sets significantly away from a precise east–west axis. And by 1714, the Julian calendar — still in use in England until 1752 — was out of synchrony with the seasons by 11 days.

Under the same parliamentary acts, Hawksmoor constructed four more churches whose axes are not oriented precisely east–west: St Alphege’s in Greenwich, St George-in-the-East in Wapping, St George in Bloomsbury, and St Mary Woolnoth in the City of London. But each of these was built on a site with physical constraints that would have made attempts at correct orientation difficult. In the case of St-George-in-the-East, Hawksmoor had petitioned to knock down adjacent houses to open up the site, but without success.

Halley was active in the building commission in the summer of 1714, attending meetings and visiting sites. The foundations for Christ Church Spitalfields, and St Anne’s, Limehouse, were laid during that summer. It seems extremely likely that Halley collaborated with Hawksmoor to assure their precise alignment with the help of his declination-corrected compass.

Ironically it has become difficult to determine the direction of east in some parts of London today. Magnetic noise, from the London Underground for example, can lead to significant local distortions of the geomagnetic field.

Heike Langenberg is a physical sciences editor at Nature.

Not content with overturning the public’s view of science, Watson has also had a major influence on science textbooks. His *Molecular Biology of the Gene* established a new style for textbooks — using concepts as crossheads, for example — which has been much copied. Gavin Borden, publisher of the 63-volume *James Joyce Archive*, was keen to publish college biology textbooks and approached Watson. They began assembling authors, who gathered at Watson’s home at

Martha’s Vineyard in the summer of 1978, and *Molecular Biology of the Cell* was born. Experts who reviewed chapters were sure it was too difficult to be an undergraduate text, but Watson was convinced that it was just what was needed — and he was, of course, right. It is both beautiful and enormously successful.

The Writing Life of James D. Watson provides valuable insights into the process that led to this success. His contribution to

solving the structure of DNA was highly significant, but if he and Crick had not worked out the structure, Franklin and Aaron Klug would have done so soon afterwards. Watson himself regards his writing, which could not have been done by anyone else, as an even greater achievement than his work on DNA that led to a Nobel prize.

Lewis Wolpert is at the Anatomy Building, University College London, Gower Street, London WC1E 6BT, UK.

The rise of the professional

The Victorian Scientist: The Growth of a Profession

by Jack Meadows

British Library: 2004. 192 pp. £16.95, \$35

John Waller

In 1873 the Victorian scientific polymath Francis Galton sent out a printed questionnaire to the most distinguished fellows of the Royal Society in London. His aim was to work out the ingredients of heredity, education, outlook and aptitude that went into making a good scientist. The results were published in his 1874 book *English Men of Science*. To nobody's great surprise, this dogmatic hereditarian concluded that most scientists were born rather than made. Although he did assert that a religious education is harmful to the development of a free, enquiring spirit, he focused almost exclusively on the personal traits of the scientists, not on the social structures and contexts that impeded or aided their research. And yet, as Jack Meadows explains in *The Victorian Scientist*, Galton was living through, and even helping to bring about, a revolution in the way science was done.

British science, it is generally acknowledged, finally became a professional activity during the lengthy reign of Queen Victoria. It was a calling once dominated by amateurs, many of them clergymen with undemanding livings, barristers with few briefs or little interest in the law, and gentlemen of leisure unfulfilled by rural sports. But by 1900 it had emerged as a full-time occupation practised by salaried experts. It is also widely agreed that this was a virtually unqualified boon. The professionalization of science, which began in German universities but later spread across the developed world, was a vital prerequisite for the discoveries that have made science such a powerful force in the modern world.

In an accessible style, Meadows charts the long struggle that saw science enter school and university curricula despite fulsome opposition from the admirers of ancient languages. He explores how scientists forged links with both industry and universities, and demonstrated the practical advantages of scientific discovery (not least Lord Kelvin's contribution to laying the undersea cables for international telegraphy), as well as the potential of pure research. Meadows discusses the transfer of mainstream science from the garden, field and makeshift laboratory to the purpose-built and lavishly equipped labs of the late nineteenth century.

This book also shows how the prestige of science rose, and with it the willingness of government to bankroll its endeavours. Less



Changing rooms: laboratories were transformed when professional scientists replaced amateurs.

and less often were scientists of humble means obliged to eke out a living writing textbooks and going on lecture tours. As the state's purse strings loosened, the social complexion of science altered, with more members of the middle and lower classes of society entering its ranks. But, as Meadows points out, there were plenty of losers in this process of professionalization, as those with dual loyalties to science and other callings were increasingly forced out, demoted to the rank of mere collectors and made to feel unwelcome at the 'high tables' of science, such as the Royal Society.

Meadows adds colour to this story with plenty of striking anecdotes, such as William Buckland dining on bluebottles, and Lyon Playfair's near-suicidal chemistry experiments. There are mentions too of Lord Kelvin's self-belief and Thomas Huxley's wit. But Meadows also makes more serious points. There is much more to telling the history of science than cataloguing its intellectual attainments. And, more importantly, the achievements of science owe at least as much to salaries and state support as to flashes of brilliant insight.

The received model of scientific progress — that it was brought about by a handful of other-worldly "scientific Shakespeares", to use C. P. Snow's phrase — is neither realistic nor especially helpful. Charles Darwin seldom called on the government for financial assistance, but his great breakthrough was critically dependent on Britain's trading interests, a well-funded navy, and a vast network of trading posts, consulates and scientific societies back home that encouraged his

activities and helped him make sense of his finds. One somewhat embittered contemporary pointed out further advantages that Darwin enjoyed. "Darwin is an enviable man," he wrote, having "a pleasant place, a nice wife, a nice family, station neither too high nor too low, a good moderate fortune, and the command of his time." These last three were particularly important. As Meadows emphasizes, the quality and quantity of science rose dramatically when, having been put on the state payroll, scientists were seldom hamstrung by having poorly connected relations or little inherited wealth.

One serious omission from this book is any real discussion of what the term 'professional' means, then and now. For instance, some Victorian scientists thought the word implied creating not university posts but a powerful intellectual élite made up largely of independently wealthy gentlemen. Nor does Meadows consider how professions operate as technocratic monopolies, often hostile to the non-professional, no matter how competent his or her research. Huxley and his allies, for example, fought hard to exclude amateurs from Victorian science, even though many of the greatest discoveries of the day had arisen from the earnest endeavours of those who, like Darwin, had neither laboratories nor official positions.

Nevertheless, *The Victorian Scientist* is to be recommended for giving non-specialist readers a much more complete picture of the context of nineteenth-century science. ■

John C. Waller is in the Department of History and Philosophy of Science, University of Melbourne, Victoria 3010, Australia.

Schrödinger's mousetrap

Part 5: Refracted glory.

Remco Zegers

Despite the fact that she was going to be interviewed by the police, Petra Pruszczyński couldn't resist a little smile as Lister entered the room. She had just caught him practising her name in the hallway and he now entered with a blush. "Miss ... Professor Prus-zinki, please sit down," he said. Obviously, the practice hadn't helped.

"Pleaz. Call me Petra," she said, and could see the relief on Lister's face. This was not going to be as difficult as she had feared it would be.

"Well then ... Petra. I have some important questions for you concerning the death of Rufus Jaeger, so let me get straight to the point. How would you describe the relationship between you and Professor Jaeger?"

Pruszczyński smiled again. "Iz no zecret that I not like him. He doez ... did bad physics and made bad image of uz all."

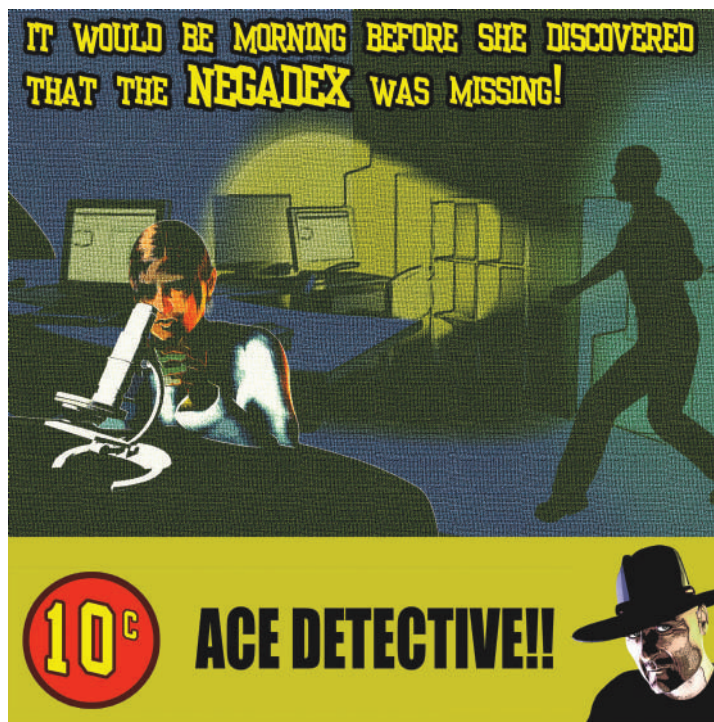
"Wouldn't you say it goes a bit further than that? I don't know about etiquette between physicists, but to accuse a person of stupidity during a major presentation is a bit harsh, isn't it?" Lister stared her straight in the eyes, and she took a moment before replying.

"Ze truth iz not always kind, az I am sure you know. I told Rufus before hiz methods unprofzessional. He never listen. I don't understand why Fenton ask him to give lecture. When I visit Fenton a few months ago he agreed with me that Rufus iz no good."

"If the truth is what you are interested in ... Petra," stressing her name, "why would you hire a person like Mr Feng, who has a rather questionable record in terms of truthfulness and was trained by this 'bad physicist'?"

Pruszczyński was somewhat surprised that the police had already figured this out, but this thought was overwhelmed by the irritation she felt at the use of her first name in this manner. "Jirong iz exzellent researcher and cannot be blamed for Rufus's fail-urez. I zee talent, I get it," and she stared back at Lister with a look that could only be interpreted as seeing a lack of talent right there.

Lister didn't seem to care and, after thinking for



a moment, asked in a much milder tone: "What are you and Mr Feng working on now?"

Pruszczyński raised her eyebrows. "We study material with very zpecial optical properties. Iz rather technical."

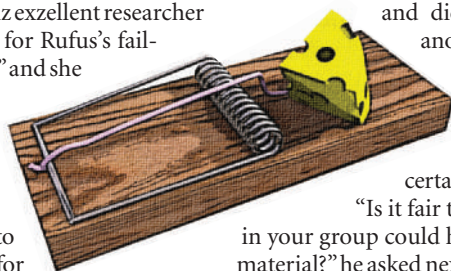
"Are you referring to ... er ... negadex, which has a negative index of ... um ... refraction?" Lister asked, trying to act nonchalantly while reading the technical terms from his notes.

Suddenly, Pruszczyński was alarmed. How could this person know about her discovery? Someone must have had difficulty keeping his mouth shut. *Psiakoś!* "Yes," she answered, trying rather unsuccessfully to hide her emotions.

"Interesting, interesting..." Lister said, seemingly pretending to know what he was talking about. "Could you make this negadex look just like an ordinary prism?"

Pruszczyński very much felt like asking what an ordinary prism looks like to put Lister in his place, but she suddenly understood what he was getting at and didn't get further than another "Yes". She wished that Lister would stop staring at her. He might be stupid, but his piercing gaze was certainly discomfiting.

"Is it fair to say that only people in your group could have had access to this material?" he asked next.



"Yes, that iz correct," she answered quickly before realizing that he had led her into a trap. How could she have been so stupid? Then she remembered the break-in at the lab a few months ago. Some pieces of negadex had been stolen along with computers and expensive optical equipment. It was certainly good to mention the break-in to keep the heat off her group. "But we had break-in lazt zummer. Zome negadex got stolen also." She wasn't sure whether Lister would recognize her hesitation. But he just responded with another "interesting, interesting".

"No one was ever caught?" he asked.

"No, waz mystery."

"Coming back to today," Lister said. "Where were you during the coffee break?"

"I discuzzed with Veronique Dubois our collaboration on a project," Pruszczyński responded.

"So, she can confirm this?" he asked.

Pruszczyński couldn't help rolling her eyes upward before stating: "One cannot zay physicists are known for lack of memory ... offizer," and felt she was back in control.

Lister didn't react though, and let his eyes drift away from Pruszczyński for a moment. He then focused his gaze back on her and said: "I think we are almost finished, but what I really don't understand, Petra, is why you are here today. You say the presentation is a farce, you obviously didn't like Rufus Jaeger and yet ... I can't believe you came here just to insult him in front of the rest of the physics community."

Pruszczyński stared right back at him. "Well, offizer, we all have to do thingz we don't want to do." And she simply couldn't resist saying: "And right now, I really don't want to be here any more. I hope we finized." She didn't even wait for Lister's nod, and left the room. All she wanted to do was to get back to her laboratory and continue her research.

To be continued...

Remco Zegers is in the National Superconducting Cyclotron Laboratory, the Joint Institute for Nuclear Astrophysics and the Department of Physics and Astronomy, Michigan State University, Michigan 48824-1321, USA.

Missed an episode? Catch the story so far at www.nature.com/news/mousetrap

CHRISTIAN DARKIN

CHRISTIAN DARKIN

Silicon shines on

Jerome Faist

Researchers are getting better at making silicon do what it really does not like to do — emit light. A silicon laser is now demonstrated that has promising features for future practical applications.

Nonlinear optics could be called the optical equivalent of the philosopher's stone: just as lead could be turned into gold by changing the number of protons in its atoms, so the colour of a laser beam can be changed from blue to red by crossing a nonlinear crystal. In nonlinear optics, incident light is converted to light of a different wavelength by making use of specific, nonlinear properties of a material. Recently, nonlinear optics has been found to be capable of performing another much sought-after trick — transforming silicon, the main material for electronics, into an optically active material. On page 725 of this issue, Rong *et al.*¹ take a further step towards a practical implementation of silicon optics by building a silicon laser that operates in a stable, continuous mode.

In the same way that steel is the base material for large-scale constructions and carbon for all known life forms, silicon is the mainstream material of electronics. It exhibits the right electronic and mechanical properties, is cheap and abundant, and can be easily processed into high-quality micrometre-scale devices. One thing that silicon could not do until recently was generate light efficiently, because of the nature of its electronic states. For this reason, all active optoelectronic applications, such as lasers for optical recording or for telecommunications, are based on group III–V materials such as GaAs and InP (ref. 2). However, as the clock frequencies of computer processors continue to increase, there is a growing need for optical data transmission that is integrated within silicon chips. Clock signals, which are needed to synchronize functions on a chip, are generated and transmitted electronically, but this scheme will run into problems with power consumption and accuracy at higher processing speeds. Optical, rather than electronic, clock distribution is expected to circumvent these problems, and the achievement of practical optical amplification in silicon would therefore be a significant advance.

The nonlinear optical effect that is used to induce light emission and amplification (laser action) in silicon is 'stimulated Raman scattering' (Fig. 1a). A laser 'pump' activates the process; a photon at the pump energy is absorbed and then re-emitted with lower energy (and so a longer wavelength) together with a 'phonon' — an elementary vibration

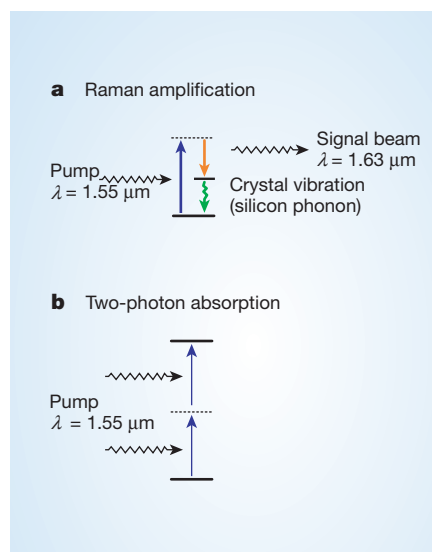


Figure 1 Raman amplification. a, In this nonlinear optical scheme, a pump photon is absorbed and re-emitted as a signal photon with a longer wavelength, along with a phonon. The process converts the pump energy into the signal beam, which is then amplified. b, Two-photon absorption, a nonlinear optical parasitic effect. It creates unwanted pairs of electrons and holes that can turn off the Raman amplification.

of the crystal³. The emitted photons make up the signal beam. By a trick of quantum wizardry, the upper energy level (dashed line in Fig. 1a) may remain virtual so that no real optical absorption is needed and the silicon crystal remains transparent. Laser action occurs because the process of light emission is stimulated — boosted — by the presence of a signal beam photon. The result is that the energy from the pump laser is transferred to the signal beam, which is then amplified.

'Raman amplification' is a small effect, and to build a laser with it you need a very high pump intensity and very low absorption losses. Such conditions have already been achieved in optical devices made from silica (SiO_2) (refs 4, 5). To achieve a sufficiently large optical intensity to produce the Raman effect in silicon, Rong *et al.*¹ used a recently developed silicon technology called silicon-on-insulator⁶. In this approach, which was originally designed to reduce the power consumption in portable electronics, thin layers of crystalline silicon with a large refractive index ($n = 3.6$) are deposited on

silica layers with a low refractive index ($n = 1.5$). The large step in refractive index enables a tight confinement of light, which can be exploited to achieve significant Raman amplification in silicon.

Using this approach, Rong *et al.* built a silicon waveguide structure, in the shape of a ridge, surrounded by silica to guide the light with low losses (Fig. 2, overleaf). They show that a pump laser with a power of only a fraction of a watt focused into this silicon waveguide creates an optical intensity up to 25 MW cm^{-2} , which is larger than what has been achieved within high-power semiconductor lasers. The Raman amplification at this intensity is still small (a few decibels per centimetre, compared with 200 dB cm^{-1} in standard semiconductor lasers) but is enough to produce laser action, owing to low optical losses in the silicon waveguide.

Raman amplification has already been shown in similar silicon structures, but the amplification was limited to very short pulses of a few nanoseconds at most^{7,8}. The problem is that an unwanted nonlinear optical side effect — two-photon absorption (Fig. 1b) — creates pairs of electrons and holes that remain for a long time in the sample and absorb both the pump light and the signal light, and so quickly turn off the Raman amplification. Rong *et al.* solve this problem by embedding the silicon waveguide within a semiconductor device, a reverse-biased p-i-n junction diode (Fig. 2). This device is designed to extract electrons and holes away from the waveguide. It is formed by implanting a short region on each side of the waveguide with impurities that convert silicon into a material with electron (n-side) or hole (p-side) conduction. A positive voltage is then applied to the n-side with respect to the p-side. In this reverse-biased scheme no current flows, but a strong electric field is generated that quickly removes the electrons and holes created by the two-photon absorption effect.

With this design Rong *et al.* demonstrate a silicon laser with continuous operation, a significant advance for the development of practical silicon lasers. Of course, the use of a nonlinear optics phenomenon means that optical pumping will always be required. However, the technique converts only a small fraction of the pump power to heat in the chip, in contrast to optically pumped lasers

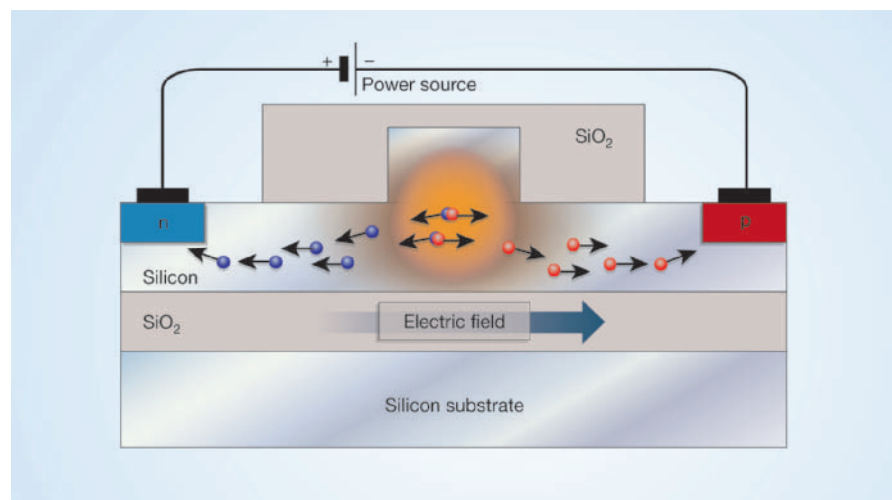


Figure 2 Cross-section of the silicon laser designed by Rong *et al.*¹. A ridge-shaped waveguide made of silicon is surrounded by silica (SiO₂). The large difference in refractive index between silicon and silica ensures that the light intensity is tightly confined within the waveguide so that a large Raman amplification can be obtained. This structure is embedded within a semiconductor device, which enhances the laser output by draining off unwanted electrons and holes that are created by the two-photon absorption shown in Fig. 1b.

that do not rely on nonlinear optical effects. This is an important advantage given that heat dissipation is becoming the key limiting parameter in microelectronics.

A fascinating feature of this work is the use of the p-i-n junction, which combines the nonlinear-optical and semiconducting properties of silicon in the same device. Rong *et al.* show that this design enables control of the optical power emitted by the laser, which in principle should also be possible at a very high frequency and could therefore be used for information processing. Last but not least, this work demonstrates that technological advances in microelectronics, in this case the silicon-on-insulator and nanolithography techniques used to fabricate the

waveguide ridge structure, can be applied to create advances in an apparently unrelated research field such as optoelectronics. ■

Jerome Faist is at the Physics Institute, University of Neuchâtel, CH-2000 Neuchâtel, Switzerland.
e-mail: jerome.faist@unine.ch

1. Rong, H. *et al.* *Nature* **433**, 725–728 (2005).
2. Rosencher, E. & Vinter, B. *Optoelectronics* (Cambridge Univ. Press, 2002).
3. Shen, Y. R. *The Principles of Nonlinear Optics* (Wiley, New York, 1984).
4. Spillane, S. M., Kippenberg, T. J. & Vahala, K. J. *Nature* **415**, 621–623 (2002).
5. Kippenberg, T. J., Spillane, S. M., Armani, D. K. & Vahala, K. J. *Optics Lett.* **29**, 1224–1227 (2004).
6. Luryi, S., Xu, J. & Zaslavsky, A. (eds) *Future Trends in Microelectronics* (Wiley-IEEE, New York, 2004).
7. Boyraz, O. & Jalali, B. *Optics Express* **12**, 5269–5273 (2004).
8. Rong, H. *et al.* *Nature* **433**, 292–294 (2005).

Cell biology

Divide and conquer

Michael Hengartner

The discovery that cell death in nematode worms induces fragmentation of mitochondria reveals a new parallel to the death process in mammals, and may shed light on why mitochondria divide in death.

When mammalian cells die by the process of apoptosis, their mitochondria fragment into smaller pieces. Why these power-generating compartments should divide as the cell around them dies, and whether this fragmentation is important for the death process or simply an epiphenomenon, has so far largely remained unclear. But an answer is suggested by the paper from Conradt and colleagues on page 754 of this issue¹. The authors show that mitochondria also fragment during apoptosis in the small nematode worm *Caenorhabditis*

elegans. Moreover, experimental induction or prevention of mitochondrial fragmentation could respectively enhance or partially prevent apoptosis. These observations hint that mitochondrial fragmentation has an evolutionarily conserved, causative role in promoting apoptotic cell death.

The term apoptosis refers to a specific type of programmed cell death that occurs in all multicellular animals, from the lowly worm to the highly complex human. Apoptosis is characterized by specific morphological changes in the dying cell,

and is mediated by several protein families².

In mammals, most apoptotic cell deaths are mediated by a specific signalling pathway known as the mitochondrial pathway. As its name implies, this pathway requires the active participation of mitochondria — the organelles better known for their role in cellular respiration and the generation of the high-energy molecule ATP. In cells condemned to die, mitochondria release several dozen proteins into the cytosol, and they can then wreak havoc in the rest of the cell.

The best known of these mitochondrial expatriates — cytochrome *c* — interacts in the cytosol with the Apaf-1 protein, ultimately activating a group of proteases (protein-digesting enzymes) known as caspases³. These enzymes then cleave a selected set of target proteins, resulting in the controlled ‘implosion’ of the cell. How cytochrome *c* *et al.* manage to cross the outer lipid bilayer of the mitochondria to reach the cytosol is still hotly debated. What is clear, however, is that this release is regulated by proteins of the Bcl-2 family, many of which can bind directly to the outer mitochondrial membrane.

Recently, several groups have reported a second peculiar behaviour of mitochondria during apoptosis: not only do they release proteins, but they also fragment into smaller pieces⁴. That mitochondria can fragment is nothing new in itself — like bacteria, mitochondria divide by a process of fission, in which one long organelle is pinched in the middle to produce two shorter daughters. Unlike bacteria, mitochondria can also undergo the reverse process, and fuse together to form long filaments. Fission and fusion are tightly controlled, and are important for the proper distribution of mitochondria during cell division.

But why mitochondria should fragment during apoptosis is not clear. One possibility is that the release of mitochondrial proteins stimulates mitochondrial division. Indeed, conditions that compromise mitochondrial function have been reported to result in short, round mitochondria. An attractive alternative would be that fission is necessary (directly or indirectly) for the release of cytochrome *c*. Consistent with this idea, interfering with the fission process has been reported to delay cytochrome *c* release during apoptosis⁵. Whether this is a general phenomenon remains to be seen. Furthermore, given that mitochondrial fission occurs continuously in living cells, there must be more to the story than fission simply promoting death.

Enter *C. elegans*. Genetic studies in this species showed that most components of the apoptotic pathway have been conserved throughout evolution⁶. For example, *C. elegans* has a Bcl-2 counterpart (CED-9), an Apaf-1-like molecule (CED-4) and a caspase (CED-3). Surprisingly, however, mitochondrial proteins have so far played at best a minor role in the apoptosis saga in *C. elegans*.

Evidence for the release of mitochondrial proteins has been rather limited⁷, and there are no data to suggest that activation of CED-3 requires cytochrome *c*. Rather, biochemical experiments led to the development of a model⁸ in which CED-9 inhibits apoptosis through direct interaction with and sequestration of CED-4. Although attractive in its simplicity, this model is still incomplete. It does not explain, for example, why CED-9 would need to be localized to mitochondria in order to function.

Conradt and colleagues¹ now address this problem. Using live microscopy, the authors noticed that cells undergoing apoptosis in *C. elegans* embryos also showed fragmented mitochondria. Fragmentation was clearly caspase-independent (it still occurred in animals with mutant, inactive CED-4 or CED-3), indicating that it was a very early event in the apoptotic programme. Furthermore, fragmentation was abrogated in animals lacking the upstream protein EGL-1 (a so-called BH3-domain protein), confirming that fragmentation is an integral part of the death process. Surprisingly, both gain-of-function and loss-of-function mutations in the Bcl-2 counterpart CED-9 also blocked fragmentation. This unexpected observation implies that CED-9 has at least two distinct, and apparently

antagonizing, functions: sequestration of CED-4, which protects cells from death, and promotion of mitochondrial fission, which the authors suggest enhances apoptosis.

So how important is mitochondrial fragmentation for *C. elegans* apoptosis? To find out, Conradt and colleagues overexpressed either the wild-type form or a dominant-negative (poison) form of DRP-1 — a protein that participates in the normal fission process⁹. As has been reported by others, overexpression of the poison form prevented fragmentation. This treatment also resulted in a mild but significant increase in long-term cell survival. In contrast, overexpression of normal DRP-1 increased mitochondrial fragmentation. Using various assays, the authors conclude that this latter treatment also led to the death of at least some cells that normally would have lived. In other words, mitochondrial fragmentation is not only associated with apoptosis in *C. elegans*, but also contributes to it.

These results are clearly exciting, but a few notes of caution are warranted. First, although Conradt and colleagues make a convincing case for a causal involvement of mitochondrial fragmentation in apoptosis (see Fig. 4 on page 758), the overall effect on cell survival was rather weak. Fewer than 20% of cells could be rescued through

expression of the poison form of DRP-1; even less could be killed by overexpression of the wild-type protein. These numbers are much lower than would be observed in animals with mutant CED-3 or CED-9, respectively. Second, because of technical limitations, the authors could only infer the extent of extra cell death caused by overexpression of wild-type DRP-1. A direct quantification is thus still missing. Third, as is the case in mammals, one must remember that mitochondrial fission occurs all the time, also in cells that live. How does the cell distinguish between normal fission and pro-apoptotic fission? Is CED-9 required for both processes, or only for the latter?

The final and perhaps most intriguing question is: how can mitochondrial division contribute to apoptosis? Conradt and colleagues¹ posit that fission might promote the release from mitochondria of a cytochrome-*c*-like molecule, which would cooperate with CED-4 to activate the protein-digesting CED-3. If correct, this would imply that the apoptotic pathways in worms and mammals are much more similar than current dogma suggests. We are sure to see more mitochondria mining in the coming years.

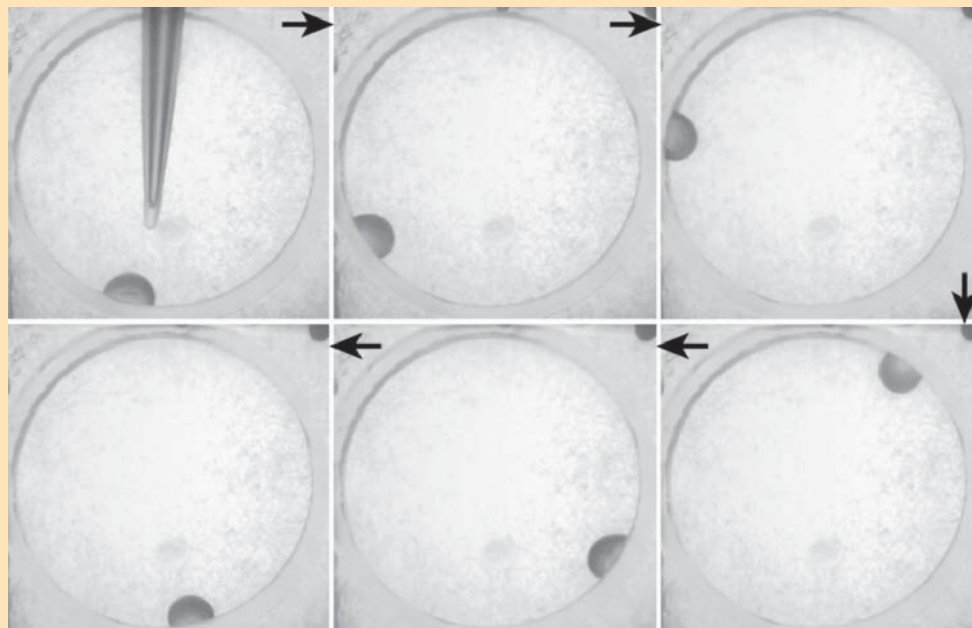
Michael Hengartner is at the Institute of Molecular Biology, University of Zurich,

Surface chemistry

Oiled acrobatics

You could be forgiven for thinking that the vision of port wine crawling up the inside of a glass was simply the result of having drunk the experiment. Yet James Thomson investigated such a phenomenon 150 years ago, and attributed the cause to gradients in surface tension. Writing in *Physical Review Letters*, Yutaka Sumino *et al.* now report that this so-called Marangoni effect makes oil droplets perform even more stunning tricks. When placed on a glass substrate submerged in water, the droplets spontaneously climb stairs or loop-the-loop on the inside of a vertical glass hoop (see pictures, in clockwise sequence from top left).

The spectacle is driven by the action of surfactant molecules, which attach uniformly to the surface of glass substrates placed in solution, thereby creating a hydrophobic coating. But when the surfactants are next to a droplet, they move from the surface into the oil phase. The resulting local perturbation of the coating makes the surface less hydrophobic, and



creates a gradient in the surface tension between the front and rear of the droplet that induces motion.

Soon after a droplet has passed by, the surfactant molecules that it has removed are replaced from the solution, returning the substrate to its hydrophobic state. Unlike in other self-running droplet systems, trajectories can therefore cross each

other repeatedly. It is this feature that makes the droplet's loop-the-loop possible. It also means that, by using narrow glass strips, the droplets can be forced from random movement into regular back-and-forth motion.

All movement stops after a few tens of seconds, however. Sumino *et al.* attribute this to the depletion

of charged complexes in the oil droplets. As long as they are present, the complexes pair up with surfactant molecules and accelerate them across the oil–water interface. Without this effect, the surface-tension gradient falls below the value needed to drive droplet motion against viscous damping.

Magdalena Helmer

Winterthurerstrasse 190, CH-8057 Zurich,
Switzerland.
e-mail: michael.hengartner@molbio.unizh.ch

- Jagasia, R., Grote, P., Westermann, B. & Conradt, B. *Nature* **433**, 754–760 (2005).
- Daniel, N. N. & Korsmeyer, S. J. *Cell* **116**, 205–219 (2004).
- Hengartner, M. O. *Nature* **407**, 770–776 (2000).

- Karbowski, M. & Youle, R. J. *Cell Death Differ.* **10**, 870–880 (2003).
- Frank, S. *et al.* *Dev. Cell* **1**, 515–525 (2001).
- Kinchen, J. M. & Hengartner, M. O. *Curr. Top. Dev. Biol.* **65**, 1–45 (2004).
- Parrish, J. *et al.* *Nature* **412**, 90–94 (2001).
- Horvitz, H. R. *Cancer Res.* **59**, 1701s–1706s (1999).
- Labrousse, A. M., Zappaterra, M. D., Rube, D. A. & van der Bliek, A. M. *Mol. Cell* **4**, 815–826 (1999).

Evolutionary biology

The power of natural selection

Andrew P. Hendry

Adaptation by natural selection is the centrepiece of biology. Yet evolutionary biologists may be deluding themselves if they think they have a good handle on the typical strength of selection in nature.

The one constant in our world is change — change often wrought by our own devices. In consequence, some of the populations and species with which we cohabit have difficulty persisting. Yet organisms should be able to adapt to changing environments, as they have done for billions of years, diversifying into a bewildering array of environments. But extinctions are also a prominent feature of the past. Were these lost organisms unable to adapt to change? If so, are the rapid changes now being driven by humans too much for adaptation to combat? At the heart of these questions is the power of natural selection to bring about evolutionary adaptation in natural populations. Writing in *Evolution*, Joe Hereford and colleagues¹ bring this matter into stark relief.

The primary mechanism of adaptive evolution is natural selection, whereby organisms possessing traits that improve their evolutionary ‘fitness’ — their survival and reproduction — contribute more genes to subsequent generations. Yet perceptions of the power of selection have recently swung at the end of a pendulum. Charles Darwin felt that “natural selection will always act very slowly, often only at long intervals of time, and generally on only a very few of the inhabitants of the same region at the same time”². If Darwin was right, natural selection should be almost imperceptible, and adaptation must require “the long lapse of ages”². This perception held sway for more than a century before it was challenged by a series of empirical studies — most famously those showing dramatic changes in the coloration of peppered moths during industrialization³. These studies inspired a wave of interest in actually measuring selection and adaptation in natural populations.

By the mid-1980s, enough studies had accumulated for John Endler to profitably review them in his classic book *Natural Selection in the Wild*⁴. Reviews of this sort typically collate and combine selection estimates for a variety of traits and studies so as to address general questions about the strength

of selection (Box 1). Endler’s review heralded a shift in our perceptions when he emphasized that “strong selection is not rare and may even be common”⁴, basing this conclusion largely on the observation that some studies documented quite strong selection.

Another way to infer the power of selection is to actually measure evolutionary changes in natural populations^{5,6}. Studies taking this approach often document substantial changes over short time intervals, suggesting that natural selection does indeed have the power to drive rapid adaptation. Darwin was too modest, it seemed, about the power of his idea.

Fast-forward to 1998, when I joined a discussion group led by Joel Kingsolver at the University of Washington. This group set about analysing all studies of natural selection published since Endler’s book.

Burdened by the practical needs of graduating, I soon bowed out of the project and did not see the results until 2001. Surprisingly, it seemed that Endler’s conclusions had swung the pendulum too far back when Kingsolver *et al.* emphasized that “directional selection on most traits and in most systems is quite weak”⁷. This conclusion was largely based on the observation that most estimates of selection were non-significant and centred around zero. A particularly worrisome finding was that most studies did not have sufficient statistical power to detect typical strengths of selection^{7,8}. Perhaps the pendulum should swing all the way back to Darwin: natural selection really is weak in nature, except in exceptional situations.

Enter Hereford *et al.*¹, who argue that previous reviews did not have objective criteria by which to judge whether selection was weak or strong. They suggest that this problem can be resolved if selection estimates for individual traits are standardized to allow comparison with the expected strength of selection on fitness itself (Box 1). Selection on fitness, they argue, provides a clear benchmark for strong selection. In reviewing many of the same studies as Kingsolver *et al.*, Hereford *et al.* conclude that selection estimates are, on average, 54% as strong as selection on fitness (31% after correction for a statistical bias). In their view, these values represent “extremely strong selection overall” and “such large estimates clearly cannot be representative of selection on all traits”¹. They then consider reasons for why current estimates of selection might be biased.

These results¹ raise some perplexing questions. Principal among them is the

Box 1 Measuring selection in natural populations

On the small island of Daphne Major in the Galapagos Islands, Peter and Rosemary Grant and colleagues⁹ measured the beak size of all medium ground finches (*Geospiza fortis*, pictured) before a drought. The abundance of seeds (particularly soft seeds) decreased during the drought and finch mortality was high. When the drought ended but before reproduction started, the Grants determined the beak size of all surviving finches.

The difference in mean beak size from before to after the drought is one measure of the strength of selection (*S*). A related measure is the selection gradient, β , which



can be obtained by dividing *S* by the variance for the trait (β can also be obtained from a regression of the trait on a measure of fitness, in this case survival). Selection estimates can be standardized by dividing *S* or multiplying β by the standard deviation of the trait. β is additionally useful because it can account for correlations among traits. The

standardized strength of selection on beak depth during the drought was $S = 0.63$ and $\beta = 0.53$. That is, selection favoured large beaks because such beaks could crack the harder seeds that remained.

Endler⁴ and Kingsolver *et al.*⁷ compiled standardized *S* or β values for many studies and traits. Hereford and colleagues¹ took a similar approach, except that β values were standardized by the mean for the trait, rather than its standard deviation. Hereford *et al.* argue that the benefit of standardizing selection by the mean is that the corresponding value for fitness should be 1. **A.P.H.**

A. P. HENDRY

apparent paradox that typical studies of selection do not have the statistical power necessary^{7,8} to detect selection that appears unrealistically strong¹. Unfortunately, this paradox will not be resolved simply by accumulating more data of the same ilk, as all reviews identify problems with our current methods^{1,4,7,8}. How, then, are we to obtain a good handle on the true power of selection in nature?

Evolutionary biologists will have to resolve this uncertainty by determining how best to measure and judge the strength of selection, and by conducting more robust studies of selection. Meanwhile, we are only deluding ourselves that we have a good handle on the typical power of selection in nature. Once we do, we can begin to

investigate how humans are changing selection pressures, and whether populations and species will be able to adapt accordingly. ■

Andrew P. Hendry is in the Redpath Museum and Department of Biology, McGill University, Montreal, Quebec H3A 2K6, Canada.
e-mail: andrew.hendry@mcgill.ca

1. Hereford, J., Hansen, T. F. & Houle, D. *Evolution* **58**, 2133–2143 (2004).
2. Darwin, C. *On the Origin of Species* (John Murray, London, 1859).
3. Kettlewell, H. B. D. *The Evolution of Melanism: The Study of a Recurring Necessity* (Oxford Univ. Press, 1973).
4. Endler, J. A. *Natural Selection in the Wild* (Princeton Univ. Press, 1986).
5. Hendry, A. P. & Kinnison, M. T. *Evolution* **53**, 1637–1653 (1999).
6. Stockwell, C. A., Hendry, A. P. & Kinnison, M. T. *Trends Ecol. Evol.* **18**, 94–101 (2003).
7. Kingsolver, J. G. *et al.* *Am. Nat.* **157**, 245–261 (2001).
8. Hersch, E. I. & Phillips, P. C. *Evolution* **58**, 479–485 (2004).
9. Grant, P. R. & Grant, B. R. *Evolution* **49**, 241–251 (1995).

Planetary science

Saturn's mixed magnetosphere

Fran Bagenal

When interplanetary shock waves hit the Cassini spacecraft and then Saturn in January 2004, it presented a unique opportunity to study the planet's magnetosphere and to compare it with that of Earth.

Saturn can be considered as the geometric mean of Earth and Jupiter in terms of the strength and extent of its magnetic field. Three papers in this issue — by Clarke *et al.*¹, Kurth *et al.*² and Cray *et al.*³ — describe the response of Saturn's magnetosphere to changes in the solar wind as observed by NASA's Cassini spacecraft and the Hubble Space Telescope (HST). The authors conclude that some aspects of the behaviour of Saturn's magnetosphere are similar to the behaviour of Earth's magnetosphere, some to that of Jupiter's and some are unique. Studies of Saturn's magnetic field and how it is driven by the solar wind are interesting in their own right, but they also allow researchers to compare different planetary magnetospheres and to test our understanding of Earth's system by applying the same principles to different conditions.

Earth's magnetic field forms a cavity in the solar wind — the stream of electromagnetic radiation and charged particles that flows outwards from the Sun. Earth's magnetosphere extends roughly 10 times the planet's radius towards the Sun and many hundreds of Earth radii away from the Sun, in a 'magnetotail' stretching downstream of Earth in the solar wind. Jupiter is much larger than Earth (by a factor of 11), and its magnetosphere is also vast, extending 50–100 jovian radii on the dayside, with a magnetotail that stretches out to its orbit distance. Saturn's magnetosphere (Fig. 1, overleaf) is an intermediate case, extending about 20 Saturn radii towards the Sun (Saturn's radius is 9.4 times that of Earth).

Most of the material in Earth's magnetosphere is a plasma of protons and electrons that has leaked in from the solar wind. By contrast, the magnetospheres of Jupiter and Saturn are mainly fed by plasma sources of heavy ions from their satellites.

The three papers^{1–3}, beginning on page 717, describe observations of magnetosphere dynamics at Saturn. In Earth's magnetosphere, plasma circulates in a flow pattern that is primarily driven by the coupling of the planetary magnetic field to the solar wind. Within about 15° of the poles, Earth's magnetic field is directly connected to the solar wind. At lower latitudes the magnetic field topology is closed, with magnetic field lines connected at both ends to the planetary dynamo. At the outer boundary of the magnetosphere — the dayside 'magnetopause' — small regions of closed magnetic field couple to the solar magnetic field (which is swept towards the planet by the solar wind) in a process called magnetic reconnection. Once coupled to the solar wind, these tubes of magnetic flux are swept back over Earth's poles and down the magnetotail where they reconnect to closed field lines — as they must, to conserve the total magnetic flux from the planet.

The stresses associated with this process of coupling solar wind and magnetosphere drive electrical currents between the magnetopause and the ionosphere (the ionized upper part of the planet's atmosphere), leading to radio and auroral emissions. The terrestrial aurorae form in rings around Earth's magnetic poles, at the boundaries



100 YEARS AGO

What mutation is in biology, conversion is in psychology, and revolution in sociology. It may be said that to assume such parallels is merely to beg the question, but I think that the apparent parallelism cannot be without significance... If the supposed analogy is a valid one, it appears to follow that mutability is due to the same general causes as ordinary variability (just as change of opinion and reform are due to the same general causes as conversion and revolution), but that there is this difference — mutability represents an explosion of energy, as it were, in a given direction, and therefore differs from ordinary variation somewhat as the firing of a gun differs from the explosion of a loose heap of powder... [T]he chance of mutations succeeding from the first is comparatively remote, though such a thing is quite possible; but since they are the result of general causes, the sort of changes the mutations exhibit are likely to come about in due course, just as the sort of changes represented by a revolution are likely to prevail ultimately, though the revolution itself may appear to fail. T. D. A. Cockerell
From *Nature* 16 February 1905.

50 YEARS AGO

Amazon Head-Hunters. By Lewis Cotlow. The author of this book is a New York insurance broker whose hobby is travelling in lands inhabited by primitive races... Between 1940 and 1949 he made several expeditions to the north-west of the South American continent... These are the areas inhabited by the Choco, Colorado and Yagua Indians, and include also the very isolated country of the Jivaro Indians, who are especially known for their custom of drying and shrinking the heads of their enemies... Mr. Cotlow was able to become very friendly with several of their chiefs, and they informed him of the number of heads which they had taken during their lives. He brings out forcibly the fact that the relatives of a man slain in battle are in honour bound to kill his killer and to shrink his decapitated head. The relatives of this victim must retaliate in the same manner, so that inter-community warfare is almost continuous. The author describes fully the method of shrinking a head. Unfortunately, he did not actually see it carried out, since at the time of the raid he was stricken with dysentery.
From *Nature* 19 February 1955.

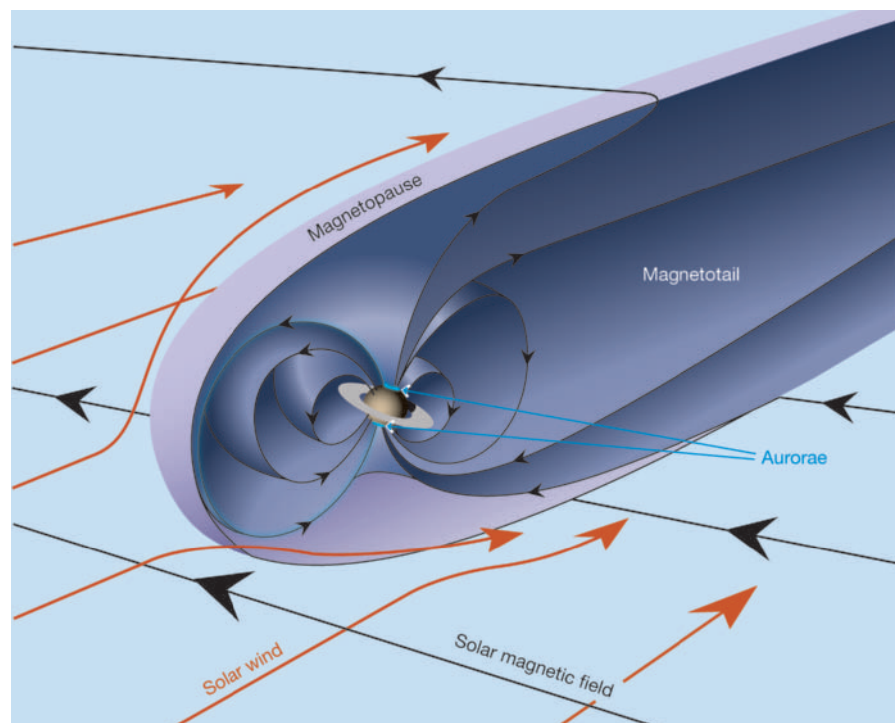


Figure 1 Saturn's magnetosphere. The planet's magnetic field forms a cavity in the solar wind that extends roughly 20 Saturn radii towards the Sun and stretches several hundreds of radii away from the Sun in a magnetotail. The size of the planet and rings are exaggerated by a factor of two. Clarke *et al.*¹, Kurth *et al.*² and Cray *et al.*³ report observations showing that, as occurs around Earth, the solar wind influences the auroral emissions and dynamics of the magnetosphere, but also that the controlling factors are different from those at Earth.

between regions of open and closed magnetic field. The orientation of the solar magnetic field is important for magnetic reconnection and therefore strongly modulates dynamical activity and auroral emissions.

Unlike the situation near Earth, Jupiter's intense aurorae are caused by stresses associated with rotation rather than coupling to the solar wind. Jupiter's strong magnetic field couples profuse plasma from its moon Io to the planet's rapid spin (a 9.9-hour period). As the plasma spreads out and inflates the giant magnetosphere, strong electrical currents between the magnetosphere and Jupiter's ionosphere keep the plasma rotating (against the tendency to slow down owing to conservation of angular momentum). Observations of Saturn in 1981 by the Voyager spacecraft indicated that its magnetosphere dynamics may be similar to those of Jupiter and be driven by Saturn's fast rotation (a 10.7-hour period). However, Saturn's icy satellites are much weaker plasma sources than Jupiter's volcanic Io, making Saturn's magnetosphere less inflated, the stresses less, and the aurorae weaker. Moreover, earlier HST studies hinted that the solar wind influences the aurorae. Thus, when Cassini approached Saturn, scientists grabbed the chance to test the hypothesis that its magnetosphere dynamics and associated aurorae are controlled by the solar wind, as occurs around Earth.

For 22 days, Cassini's instruments measured the magnetic field, plasma density and

plasma velocity in the solar wind while the HST and Cassini radio antennas monitored Saturn's auroral activity. Nature cooperated and provided a couple of interplanetary shock waves that passed the Cassini spacecraft on 15 and 25 January and hit Saturn's magnetosphere some 17 hours later. Clarke *et al.*¹ report HST observations of the subsequent brightening of auroral emission, and Kurth *et al.*² report accompanying increases in radio emission. Cray *et al.*³ show a correlation of auroral intensity with solar-wind dynamical

pressure, supporting the view that the solar wind has an Earth-like role at Saturn.

But further study showed that the solar-wind conditions that influence Saturn's magnetosphere dynamics are different from those that influence Earth's. Compression of the magnetopause by the solar wind is more important than reconnection of the solar and saturnian magnetic fields. Cray *et al.*³ point out that at Saturn's orbit (9.5 times Earth's distance from the Sun), the solar magnetic field is fairly closely confined to the plane of Saturn's orbit around the Sun; the solar and saturnian fields are therefore largely at right angles to each other, so that the optimum conditions for magnetic reconnection are not met.

But there are further unknown complications to Saturn's magnetospheric dynamics and auroral activity: what are the effects of Titan, Saturn's largest moon, and of the 26° tilt of Saturn's southern pole towards the Sun? Studies of the terrestrial magnetosphere have shown the need for long-term monitoring to properly distinguish between the causes of various auroral effects. Cassini will make many more observations in its 75 orbits of Saturn. But the imaging spectrograph onboard the HST is no longer operational and the space telescope's future looks bleak. Furthermore, these vital but unglamorous synoptic studies of planetary magnetospheres currently have little appeal for NASA, whose science budget is being squeezed to pay for human space exploration. So it may be some time before we can follow up on these latest findings from Saturn.

Fran Bagenal is in the Laboratory for Atmospheric and Space Physics and the Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, Colorado 80309-0392, USA.
e-mail: bagenal@colorado.edu

1. Clarke, J. T. *et al.* *Nature* **433**, 717–719 (2005).
2. Kurth, W. S. *et al.* *Nature* **433**, 722–725 (2005).
3. Cray, F. J. *et al.* *Nature* **433**, 720–722 (2005).

Neurodegeneration

Cellular defences destroyed

Dennis W. Choi

A lack of blood flow can kill nerve cells, by causing a massive influx of calcium ions. But what's happened to the cellular mechanisms for coping with excess calcium?

Without calcium, life would simply not be possible: it imparts strength to bones and serves as a messenger in myriad cellular processes. Underlying this ubiquitous signalling role is a 10,000-fold concentration gradient across the cell membrane, with 50–100 nM free calcium ions inside cells, separated by only a lipid bilayer from 1 mM calcium outside. Exquisitely regulated channels permit

dollops of calcium to rush in across the membrane, causing localized increases in intracellular calcium levels and the activation of appropriate molecules. Yet under pathological conditions, such as a stroke, calcium can also be a killer, flooding into neurons and inducing lethal derangements. Although much has been learned about how excess calcium gets in, it has been puzzling that the mechanisms normally

responsible for calcium extrusion are not better able to cope with the deluge. A paper by Bano *et al.*, just published in *Cell*¹, provides insight into why exactly these mechanisms may fall short, just when we need them the most.

When the brain experiences ischaemia — the blood-flow shortage associated with, for instance, a stroke — the excitatory neurotransmitter glutamate is released from nerve terminals and from other brain cells, and accumulates excessively in the extracellular space. This results in 'excitotoxic' overstimulation of the receptors found on most neurons, inducing large quantities of Ca^{2+} ions to enter these cells (Fig. 1). NMDA-type glutamate receptors, whose membrane channels have a naturally high Ca^{2+} permeability, contribute prominently to this Ca^{2+} influx and toxicity². Moreover, in some neurons subjected to ischaemia, AMPA-type glutamate receptors develop enhanced Ca^{2+} permeability and become major direct routes for toxic Ca^{2+} entry³. The problem is compounded by the release of Ca^{2+} from internal stores and by the entry of Ca^{2+} through several other types of membrane channel (see, for instance, refs 4, 5).

Working on the assumption that increased Ca^{2+} entry or release might not be the whole story, however, Bano *et al.*¹ turned a spotlight on the membrane proteins that bear the lion's share of the responsibility for maintaining low intracellular Ca^{2+} concentrations — the $\text{Na}^+/\text{Ca}^{2+}$ exchanger (NCX) family. Under normal conditions, these proteins transport Na^+ ions down their concentration gradient into cells, and harness the resulting energy to transport potentially large amounts of Ca^{2+} back out. Why, wondered Bano *et al.*, isn't NCX on the job in ischaemic neurons, at least after the restoration of blood flow and Na^+ gradients?

The authors found that NCX, in particular the NCX3 isoform, is cleaved in rat brains after a simulated stroke (induced by a 15-minute blockage of the middle cerebral artery, followed by restoration of blood flow). They linked this event back to excitotoxicity by identifying the same cleavage in cultured cerebellar granule neurons that were bathed in glutamate. The authors also identified the Ca^{2+} -dependent protein-cleaving enzyme calpain as the culprit, given both the sensitivity of these events to a pharmacological calpain inhibitor, and the size of cleavage fragments. Recognizing the limited specificity of pharmacological protease inhibitors, Bano *et al.* also overexpressed a portion of a natural calpain inhibitor, calpastatin, in granule neurons; this elegant molecular probe also blocked glutamate-induced NCX3 cleavage.

Having established an association between excitotoxic or ischaemic conditions and calpain-mediated degradation of NCX,

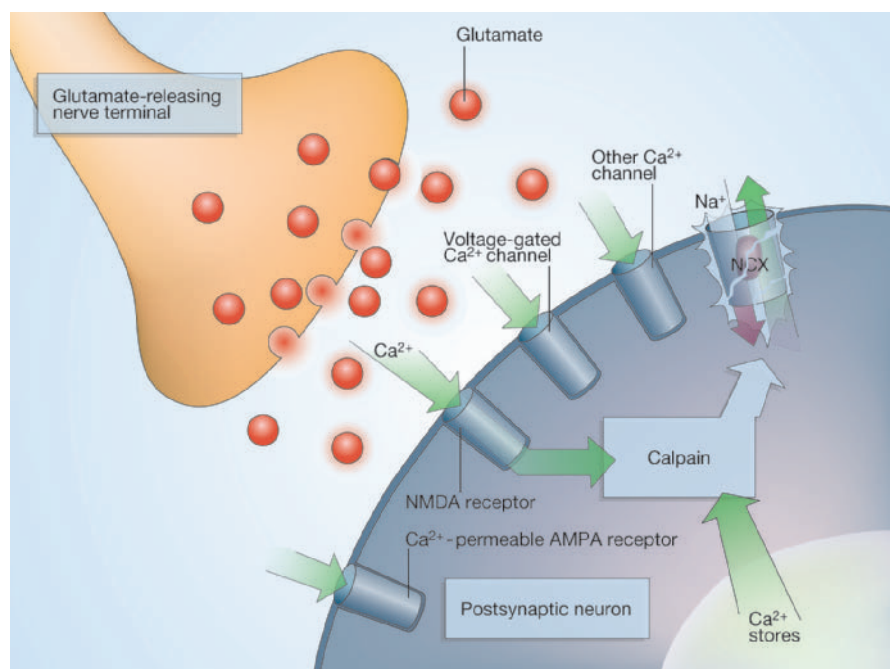


Figure 1 Glutamate overproduction and calcium flow. In response to ischaemia (a reduction in blood flow), the neurotransmitter glutamate is released from neurons and other brain cells (astrocytes; not shown). This induces Ca^{2+} entry (green arrows) into responsive neurons through NMDA-type glutamate receptors and Ca^{2+} -permeable AMPA-type glutamate receptors. Ca^{2+} also enters through voltage-gated Ca^{2+} channels and several other types of Ca^{2+} -permeable membrane channel, and is released from intracellular stores. The NCX proteins represent a primary cellular defence against Ca^{2+} overload, pumping Na^+ ions in and Ca^{2+} ions out. But Bano *et al.*¹ show that in ischaemic conditions NCX is destroyed by the Ca^{2+} -activated protease calpain.

Bano *et al.* went on to delineate its functional importance. They showed that overexpressing either calpastatin or the calpain-resistant NCX isoform, NCX2, reduced the glutamate-induced increase in both neuronal intracellular Ca^{2+} concentration and neuronal death. Finally, the authors found that suppressing NCX3 expression, achieved by treatment with small interfering RNAs, had the opposite effect: granule neurons became sensitized to glutamate-induced Ca^{2+} overload and death.

Bano and colleagues' observations¹ provide evidence that the calpain-mediated cleavage of NCX can be a 'feedforward' step in excitotoxic neuronal death. It is highly plausible that compromising the cell's main high-capacity mechanism for removing Ca^{2+} would increase the impact of the initial glutamate-induced burst of Ca^{2+} entry. More than a decade ago, Khodorov and colleagues⁶ reported that the sustained increase in the intracellular Ca^{2+} concentration to which cerebellar granule cells become victim after exposure to high glutamate levels persisted after the removal of both glutamate and extracellular Ca^{2+} . This led them to suspect a disruption of homeostatic mechanisms. Moreover, they found that removing extracellular Na^+ — expected to inhibit NCX-mediated Ca^{2+} extrusion (but also to have other effects) — did not augment cell death, leading them to propose specifically that NCX was suppressed. Bano *et al.* have now

provided solid molecular evidence to support this view.

Further studies will be needed to answer several questions that arise. In particular, given the known heterogeneity in the characteristics of excitotoxicity in different types of neuron, it will be important to see whether the timing and extent of calpain-induced NCX3 (or NCX1) cleavage varies from neuron to neuron, or from one compartment to another within the same neuron. Further, does NCX cleavage depend on the route of Ca^{2+} entry into the cytoplasm, as do other elements of the excitotoxic cascade⁷? Does NCX cleavage contribute to the death of other brain cells after ischaemia (for example, the AMPA-receptor-mediated death of oligodendrocytes, or the non-excitotoxic death of astrocytes)? Stepping back, it will be important to test whether inhibiting NCX cleavage reduces brain damage in animal stroke models. If so, this might form the basis for a new class of neuroprotective treatments, useful in strokes and perhaps other conditions.

However, one should not conclude yet that NCX cleavage always enhances neuronal death. Glutamate exposure increases intracellular Na^+ as well as Ca^{2+} levels, and, in Na^+ -loaded and depolarized cells, NCX can run in the reverse direction, mediating Na^+ efflux and Ca^{2+} influx and thus potentiating excitotoxic neuronal death⁸. NCX cleavage might be expected to oppose these events.

Furthermore, excitotoxic insults, like other insults, can induce neurons to undergo cellular suicide (programmed cell death), and intermediate increases in intracellular Ca^{2+} concentrations can stave off this death programme^{9,10}. Although NCX has low affinity for Ca^{2+} , its inhibition can prolong the transient large fluctuations in Ca^{2+} that are evoked by the depolarization of sympathetic neurons¹¹. So, one could envisage a situation in which an initial harmful burst of Ca^{2+} influx is followed by relative normalization of Ca^{2+} levels, and calpain-induced NCX cleavage helps to elevate these late Ca^{2+} levels to the point at which programmed cell death is inhibited.

As Bano *et al.*¹ point out, calpain-induced NCX cleavage has an intriguing parallel in the caspase-induced cleavage of another Ca^{2+} -extrusion pump, PMCA¹². Perhaps both of these cleavage events can either

promote or inhibit cell death in different circumstances, specifically acting at low injury levels to ensure that a commitment to undergo programmed cell death does not occur casually. ■

Dennis W. Choi is at the Merck Research Laboratories, 770 Sumneytown Pike, West Point, Pennsylvania 19486, USA.
e-mail: dennis_choi@merck.com

1. Bano, D. *et al.* *Cell* **120**, 275–285 (2005).
2. Choi, D. W. *et al.* *J. Neurosci.* **8**, 185–196 (1988).
3. Calderone, A. *et al.* *J. Neurosci.* **23**, 2112–2121 (2003).
4. Xiong, Z.-G. *et al.* *Cell* **118**, 687–698 (2004).
5. Aarts, M. *et al.* *Cell* **115**, 863–877 (2003).
6. Khodorov, B. *et al.* *FEBS Lett.* **324**, 271–273 (1993).
7. Aarts, M. *et al.* *Science* **298**, 846–850 (2002).
8. Kiedrowski, L., Czyz, A., Baranaukas, G., Li, X.-F. & Lytton, J. *J. Neurochem.* **90**, 117 (2004).
9. Franklin, J., Sanz-Rodriguez, C., Juhasz, A., Deckwerth, T. L. & Johnson, E. M. *J. Neurosci.* **15**, 643–664 (1995).
10. Choi, D. W. *Trends Neurosci.* **18**, 58–60 (1995).
11. Wanaverbecq, N., Marsh, S. J., Al-Qatari, M. & Brown, D. A. *J. Physiol. (Lond.)* **550**, 83–101 (2003).
12. Schwab, B. L. *et al.* *Cell Death Differ.* **9**, 818–831 (2002).

Neurobiology

Bright blue times

Russell G. Foster

The discovery of light-sensitive neurons that can adjust our body clocks prompted a search for their light-detecting molecule. We now know the identity of this pigment — and that these cells do more than was thought.

Our in-built ability to tune our body clocks to day and night relies on a special set of light-sensitive neurons in the eye. Three new papers — two in this issue^{1,2} and one in *Science*³ — provide strong evidence that melanopsin is the pigment that allows these cells to respond to light. Yet another paper in this issue⁴ provides the first detailed description of the cells in a primate.

Until a decade or so ago, we thought we understood the workings of the vertebrate eye. The rods and cones (photoreceptors) of the outer retina detect light, with cells of the inner retina providing the initial stages of visual processing, before ganglion cells convey information to the brain via the optic nerves (Fig. 1a). But evidence for another light-sensing system in the eye — separate from the rods and cones — began to accumulate in the 1990s. This came from researchers, studying the body clock (circadian rhythms), who appreciated that the eye performs two quite different sensory tasks. Its familiar function is to collect and process light to generate an image of the world. But it also provides a measurement of environmental brightness at dawn and dusk, to align circadian time to environmental time.

Circadian biologists had a problem locating this function within the known structures of the eye, and, in an attempt to find the cells responsible, they used mouse models that lacked rods, cones or both. Remarkably, the loss of all types of known photoreceptor

had little effect on the animals' ability to tune their circadian system to light^{5,6}, but loss of the eyes abolished this ability completely. So there had to be another light sensor in the eye. Subsequently, these sensors were shown to contribute, together with the rods and cones, to the regulation of pupil constriction and of other responses to light⁷. Because these photoreceptors are associated with brightness detection, the collective term 'non-image-forming photoreceptor system' has been used to describe them; the 'image-forming' system refers to the rods and cones.

Several groups then began to investigate which neurons mediate non-image-forming responses to light — and what makes them light-sensitive. The answer to the first problem was provided by the discovery that a small subset (around 1%) of retinal ganglion cells respond to light directly^{8,9}, in part through an increase in intracellular calcium concentrations⁹ (Fig. 1b). Finding what makes these cells responsive to light has been more difficult.

The known photopigments found in animals combine an opsin protein with a vitamin-A-based light-absorbing molecule (chromophore) called 11-*cis*-retinaldehyde. The first stage of light detection involves the absorption of a photon by 11-*cis*-retinaldehyde, and the photoisomerization of this molecule to the all-*trans* state (Fig. 2). This allows the opsin to trigger a phototransduction cascade that ultimately changes the

cell's electrical activity. All opsin–vitamin-A photopigments have a characteristic absorption profile, which allows them to be identified on the basis of their spectral responses to light. The photopigment in the mouse non-image-forming photoreceptors has a maximum sensitivity in the blue part of the spectrum, at a wavelength (λ_{max}) of 479 nm (ref. 7). This photopigment was originally termed opsin photopigment 479 (OP⁴⁷⁹), but its molecular identity remained a mystery.

Melanopsin¹⁰ (also called Opn4) soon emerged as the best candidate. It is expressed in the photosensitive ganglion cells^{8,9}, and its genetic ablation attenuates circadian and pupil responses to light. Furthermore, removing melanopsin in mice lacking all functional rods and cones abolishes such responses completely¹¹. Yet although these studies showed that melanopsin is essential for photosensitive ganglion cells to respond to light, they could not explain how melanopsin works.

Melyan *et al.*¹, Qiu *et al.*² and Panda *et al.*³ have now assessed the function of melanopsin by combining its expression with physiological assays of cellular photosensitivity. All three papers show that melanopsin can confer photosensitivity on non-photosensitive cell types, and that specific forms of retinaldehyde (especially 11-*cis*-retinaldehyde) are required. In short, they show that melanopsin acts as a photopigment.

Beyond this remarkable finding, the papers differ in some of their conclusions. Qiu *et al.*² and Panda *et al.*³ show that melanopsin has a λ_{max} very close to 480 nm — so OP⁴⁷⁹ seems to be melanopsin. However, Melyan *et al.*¹ (and previously Newman *et al.*¹²) suggest that melanopsin has a λ_{max} closer to 420–430 nm. There is no obvious explanation for this difference, but it might relate to the immediate environment of the expressed photopigment. For example, pH conditions combined with differences between mouse^{2,3} and human¹ melanopsin might be responsible.

Melyan *et al.* and Panda *et al.* also provide evidence that melanopsin exhibits bistability — the ability to bind alternately to 11-*cis*- and all-*trans*-retinaldehyde, and to act as both a sensory pigment and an isomerase for photopigment regeneration (Fig. 2). But Qiu *et al.* suggest otherwise. This discrepancy might reflect the different cell types used; for example, Qiu *et al.* expressed melanopsin in cells that show endogenous retinoid metabolism.

All three groups also look at the melanopsin-evoked phototransduction cascade. There is broad consensus from these and previous⁹ studies that light will ultimately trigger the release of calcium ions inside the ganglion cells, and that this involves some kind of interaction of melanopsin with a G protein (a member of a diverse family of proteins that link receptors to signalling cascades). Although our knowledge is far from complete, this phototransduction

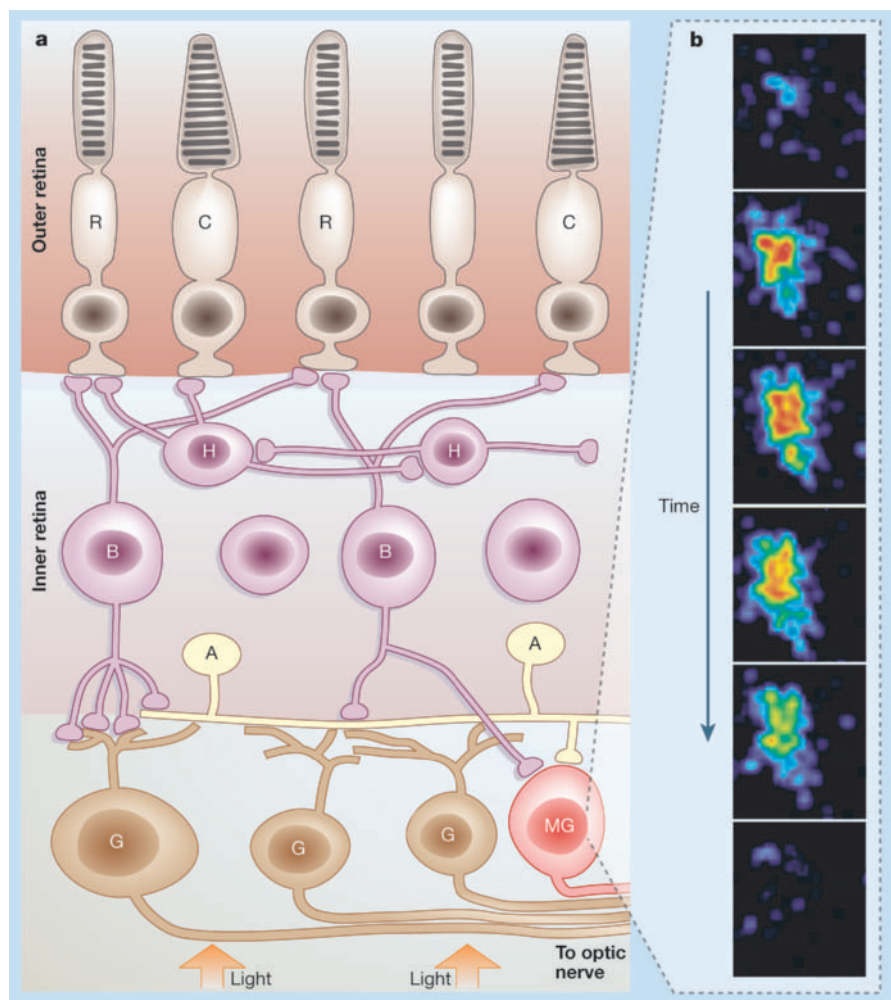


Figure 1 The basic cell types of the vertebrate retina. **a**, The rods (R) and cones (C) convey visual information to the ganglion cells (G) through the bipolar cells (B). Horizontal cells (H) allow lateral connections between rods and cones. Amacrine cells (A) allow lateral connections between bipolar and ganglion cells. The optic nerve is formed from the axons of all the ganglion cells. A subset of ganglion cells (MG cells) also detects light directly; for this, they require the photopigment melanopsin, as now confirmed^{1–3}. **b**, Light, via melanopsin, causes changes in Ca^{2+} levels in MG cells⁹ (a fluorescent Ca^{2+} indicator was used here). Counterintuitively, light passes through the transparent ganglion layer to reach the rods and cones.

cascade seems very different from that of rods and cones. In fact, it is more like that of an invertebrate photoreceptor, raising interesting questions about the evolution of ganglion-cell photosensitivity.

And what of the fourth new paper⁴? Behavioural studies in humans indicate that we possess a non-image-forming photoreceptor system like that of rodents. Dacey *et al.*⁴ have now examined the photosensitivity of melanopsin-expressing ganglion cells in macaques — a monkey with a very similar visual system to ours. The authors find that, in general terms, the responses of these cells are like those of rodents, again showing a spectral response with a λ_{max} near 480 nm. This begs the question of why blue light is so important. We can only guess at the answer, but perhaps it is no coincidence that 480-nm light dominates the wavelengths at dawn and dusk. Could it be that the main role of these ganglion cells is to detect twilight?

Previous studies in rodents have also shown that the image-forming and non-image-forming systems interact^{8,11}; Dacey *et al.* extend our understanding of these interactions. Perhaps their most fascinating finding is that the short-wavelength-detecting cones attenuate the light responses of melanopsin-expressing ganglion cells, whereas the rods and medium- and long-wavelength cones provide an excitatory input. Again, an ecological explanation remains unclear. Dacey *et al.* also show that these ganglion cells project to the lateral geniculate nuclei — the brain structure that relays image-forming information to the visual cortex. This observation, along with other recent work¹³, supports the idea that the non-image-forming system contributes to aspects of visual perception.

This final observation completes a circle of thought. Originally, rods and cones were assumed to be the only photoreceptors of the

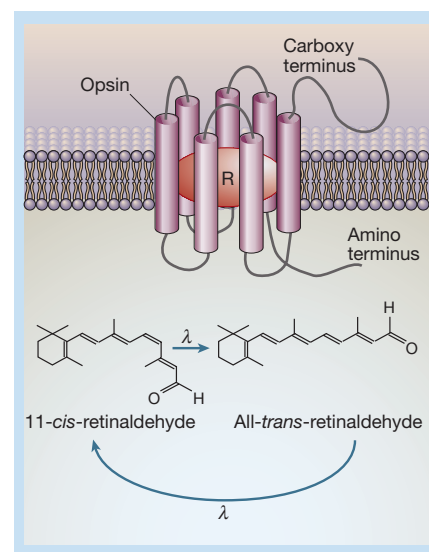


Figure 2 Animal photopigments consist of an opsin protein coupled to the 11-*cis* form of retinaldehyde (R). In response to light of an appropriate wavelength (λ), 11-*cis*-retinaldehyde absorbs a photon and is photoisomerized to all-*trans*-retinaldehyde. This changes the opsin's conformation, initiating a phototransduction cascade that includes Ca^{2+} changes in light-sensitive ganglion cells. Some invertebrate opsins, and possibly melanopsin, can act as both a photosensor and a photoisomerase — driving all-*trans*-retinaldehyde back to the 11-*cis* configuration.

eye. Then it was discovered that the loss of rods and cones had little effect on circadian responses to light, suggesting that the eyes use different photoreceptor systems for these different sensory tasks. Next, pupil constriction and circadian responses to light were shown to arise from an interaction between the two receptor systems. Finally, it now seems likely that photosensitive ganglion cells impinge directly upon image formation. Clearly, all future experiments on human light detection will have to consider the relative contributions of both photoreceptor systems. ■

Russell G. Foster is in the Division of Neuroscience and Mental Health, Charing Cross Hospital, Faculty of Medicine, Imperial College London, London W6 8RF, UK.

e-mail: r.foster@imperial.ac.uk

1. Melyan, Z., Tarrant, E. E., Bellingham, J., Lucas, R. J. & Hankins, M. W. *Nature* **433**, 741–745 (2005).
2. Qiu, X. *et al.* *Nature* **433**, 745–749 (2005).
3. Panda, S. *et al.* *Science* **307**, 600–604 (2005).
4. Dacey, D. M. *et al.* *Nature* **433**, 749–754 (2005).
5. Freedman, M. S. *et al.* *Science* **284**, 502–504 (1999).
6. Lucas, R. J., Freedman, M. S., Muñoz, M., García-Fernández, J. M. & Foster, R. G. *Science* **284**, 505–507 (1999).
7. Lucas, R. J., Douglas, R. H. & Foster, R. G. *Nature Neurosci.* **4**, 621–626 (2001).
8. Berson, D. M., Dunn, F. A. & Takao, M. *Science* **295**, 1070–1073 (2002).
9. Sekaran, S., Foster, R. G., Lucas, R. J. & Hankins, M. W. *Curr. Biol.* **13**, 1290–1298 (2003).
10. Provencio, I. *et al.* *J. Neurosci.* **20**, 600–605 (2000).
11. Hattar, S. *et al.* *Nature* **424**, 75–81 (2003).
12. Newman, L. A., Walker, M. T., Brown, R. L., Cronin, T. W. & Robinson, P. R. *Biochemistry* **42**, 12734–12738 (2003).
13. Perrin, F. *et al.* *Curr. Biol.* **14**, 1842–1846 (2004).

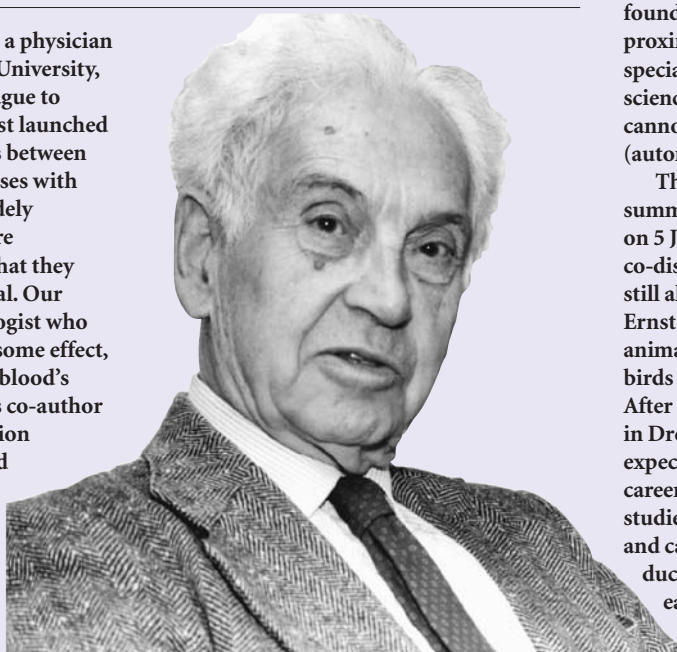
Obituary

Ernst Mayr (1904–2005)

One Sunday in 1953, my father, a physician and haematologist at Harvard University, invited a newly recruited colleague to lunch at our house. Dad had just launched a study of possible associations between human blood groups and diseases with him. At that time, scientists widely assumed that blood groups were 'selectively neutral' — that is, that they had no effect on human survival. Our guest was an evolutionary biologist who suspected that they must have some effect, perhaps one far removed from blood's familiar functions. Dad and his co-author went on to discover an association between ABO blood groups and stomach cancer, one of the first studies to show that blood groups are indeed influenced by natural selection. The co-author was Ernst Mayr, widely regarded as the greatest evolutionary biologist of the twentieth century, who died on 3 February 2005.

When I met Mayr that Sunday, I was a 16-year-old schoolboy. He later inspired me to launch a second career, parallel to my work as a membrane physiologist, on the evolutionary biology of New Guinea birds, his own early speciality. For 30 years he and I collaborated on analysing a mammoth database that he had accumulated on the distributions of island birds. The result was a co-authored 556-page book published soon after his 97th birthday. That Sunday lunch and its consequences illustrate many keys to Mayr's greatness: his capacity for close friendships and collaborations with younger scientists as well as with peers; his broad perspective that let him recognize new significance in the work of many specialists; and his capacity for sustained hard work and complex analysis.

The achievements for which Mayr is best known fall into six areas. First, as an ornithologist he was the leading expert on birds of New Guinea and the tropical southwest Pacific; he described more species and subspecies of living birds than anyone else of his or subsequent generations. Second, as a systematist he was a principal architect of what is termed the 'evolutionary synthesis', which finally succeeded in showing how the adaptive changes that natural selection produces in single populations result in the evolution of biodiversity. That synthesis fused the hitherto separate research programmes of geneticists and field naturalists, and



Evolutionary biologist, ornithologist and philosopher

explained how evolution has given rise to organisms ranging from microscopic bacteria to redwood trees.

In that process lies Mayr's third major achievement, his assembly of overwhelming evidence that most species are not collections of individuals arbitrarily delineated by taxonomists but real entities: "a group of actually or potentially interbreeding natural populations reproductively isolated from other such populations", to quote his widely cited formulation. He also demonstrated that new species of birds and mammals arise through allopatric speciation (geographical isolation of initially conspecific populations) — thereby in effect solving the problem of the origin of species that had eluded Charles Darwin despite the title of Darwin's great book.

Fourth, as a biogeographer Mayr's studies of a fauna's composition, origins, history and boundaries have served subsequent biogeographers as models for testing their own results. Fifth, as an evolutionary biologist, Mayr traced in detail the combined operation of population genetics and evolutionary processes in diverse phenomena throughout the animal kingdom, as illustrated by the study of blood groups that resulted in my meeting him.

Finally, as a historian and philosopher

of science, in recent decades Mayr clarified the regularly misunderstood central concepts of biology: teleology; the foundations of biological classification; proximate and ultimate causation; the special problems posed by historical sciences to which experimental methods cannot be applied; and the distinctness (autonomy) of biology as a science.

The facts of Mayr's career can be briefly summarized. He was born in Germany on 5 July 1904, at a time when evolution's co-discoverer, Alfred Russel Wallace, was still alive. On weekends, his parents took Ernst and his brothers on walks to observe animals, plants and fossils, among which birds in particular kindled his interest. After a rigorous high-school education in Dresden, Ernst obeyed Mayr family expectations by preparing for a medical career and completed his preclinical studies in 1925. However, his observation and careful description of a pair of a rare duck, last recorded in Germany 77 years earlier, led to his introduction to the Berlin ornithologist Erwin Stresemann. Recognizing Mayr's

talent, and also his thirst to visit the tropics, Stresemann offered Mayr two irresistible enticements: a position in the Berlin Museum, and prospects of a bird-collecting trip to the tropics, if Mayr could complete an entire PhD programme within 16 months.

Mayr accepted the challenge, worked 16 to 18 hours a day to receive his PhD in 1926, and took up the promised museum position. In 1928 Stresemann, now armed with money from Lord Rothschild and from the American Museum of Natural History (AMNH) in New York, delivered on his second promise by sending Mayr to the southwest Pacific for more than two years. The instructions given to Mayr were to explore five New Guinea mountain ranges, to solve the long-standing mystery of New Guinea's apparently rarest birds of paradise (he did, and they proved to be hybrids), and to collect birds on islands in the Solomon group that had been considered too dangerous to visit by previous collectors. Mayr succeeded beyond everyone's expectations. Having re-explored six of those mountain ranges and islands between 1974 and 2004, under the less-threatening conditions of the late twentieth century, I can testify that they are physically gruelling even today. Mayr managed to amass comprehensive bird collections there from 1928 to 1930, despite the perils of diseases, capsized canoes, forced descents of waterfalls and periodic threats of natives to kill him.

Soon after his return from New Guinea, in 1931 Mayr was appointed by the AMNH

to curate the museum's overflowing collections of Pacific island birds. For the next decade, all of his publications were technical taxonomic studies of birds, giving few signs of his broader interests, until the publication in 1942 of his first book, *Systematics and the Origin of Species*, which completed the evolutionary synthesis.

In 1953, a desire for contact with students and for wider intellectual horizons led Mayr to move to Harvard, as Agassiz Professor of Zoology, where he also served as director of the Museum of Comparative Zoology from 1961 to 1970. Following his official retirement from Harvard in 1975, he continued to publish with undiminished productivity. Fourteen of his 25 books were published after the age of 65, including one of his most important ones, *The Growth of Biological Thought*, which appeared when he was 78. He celebrated his 100th birthday with the publication of *What Makes Biology Unique?* in 2004. Whenever I talked to him during his nineties, I would ask him: "How many books are you working on now?" The answer was never less than two nor more than four.

What accounted for Mayr's remarkable originality and productivity? I came to realize that there wasn't a single explanation but the combination of a dozen of them — cognitive, organizational, emotional and social. Among the cognitive ones, he had an outstanding memory. When, in 1965, 24 years after the peak of Mayr's work on New Guinea birds, John Terborgh and I asked him to identify the stuffed bird specimens that we had just collected in New Guinea, we saw that, for each of the 1,400 species and subspecies of birds that he had discussed in his 1941 *Checklist of New Guinea Birds*, Mayr still remembered who had described it — and when and in what journal, its differences from its relatives, and its alternative names. To that memory for facts were allied outstanding visual recall (for example, he was alert to slight subspecific differences between bird specimens seen at different times in different museums) and auditory recall (the ethologist Klaus Immelmann related how, while he and Mayr were sitting on a garden bench in Germany in the 1970s, Mayr correctly identified a brief call note of an unseen bird as a grey wagtail, which he had not encountered since leaving Germany 40 years previously).

Mayr was also a quick learner: in the month before he reached New Guinea in 1928, he learned to speak Malay and Neo-Melanesian, to shoot a gun, and to skin and stuff birds. Like Darwin, he was a constantly curious field observer; also like Darwin, his wide interests let him



Ernst Mayr in the field: Mayr in 1928, with (left) his assistant, Sario, following two months of surveying birds in the mountains of New Guinea.

reinterpret the work of specialists, as he did with my father's data on blood groups. In my own collaboration with him, I was struck by his comfort with complexity: unlike many other scientists, he did not force facts into a one-factor explanation, but acknowledged the possibility of different multi-factor outcomes (such as different evolutionary trajectories for different bird populations).

Mayr himself spoke of his *Sitzfleisch* or capacity to stick to a job, just as the composer J. S. Bach attributed his prodigious musical output to mere *Fleiß* (industriousness). During Mayr's years as a museum director at Harvard, a job that absorbed his daytime hours, he maintained his scientific output by writing each morning from 4:30 until 7:30 a.m., then spending the evening reading. In the 16 months that it took him to complete his PhD by age 21, he took all of the required courses in zoology, learned and passed an exam in botany, completed a minor in philosophy, and researched and wrote his thesis. When he arrived at the AMNH on 20 January 1931, he was given a one-year appointment with the understanding that reappointment would depend on productivity. He published his first paper two months later (a reclassification of kingfisher subspecies based on measuring hundreds of specimens), and finished 11 more papers by the year's end. (That convinced the AMNH to renew his appointment.)

Despite not visiting an English-speaking country until his twenties, Mayr mastered English as a second language to

the point where his English prose style was widely admired for its clarity. In addition to publishing 25 books and more than 700 papers and directing Harvard's museum for nine years, he designed the AMNH's bird exhibit hall, integrated the Rothschild collection of 280,000 bird specimens into the AMNH collection, and edited the last eight volumes of the *Checklist of the Birds of the World* (a critical reassessment of all bird taxa down to the subspecies level). Each of those 'additional' achievements was a mammoth undertaking in itself.

Mayr was self-confident without being overconfident. And he could change strongly held views when presented with new evidence, as when he abandoned his initially lamarckian belief in the inheritance of acquired characteristics. His confidence in his abilities included recognition of their limitations: for instance, he resisted friends' suggestions that he expand his 1963 book *Animal Species and Evolution* to include plants and microorganisms, because of his insufficient familiarity with them. Those limitations also involved mathematics beyond algebra, which he did not use. He maintained a low opinion of the value of the cladistic methods now dominant among taxonomists. That contributed to the distance, in his later years, between his views and those of some other evolutionary biologists now active. They felt that his work belonged to the past; he felt with exasperation that they ignored much of the knowledge already gained in the past.

A widespread misconception is that great scientists tend to be loners. Actually, outstanding success in most areas of science requires outstanding social skills, as illustrated by Mayr's relationships with a wide variety of people. He achieved such good understanding with New Guinea and Solomon tribespeople in the 1920s that they not only led him in and out of areas where other Europeans feared being killed, but they also taught him their local names for birds and brought him hundreds of specimens of bird species missed by other European collectors. He once explained to me that a secret of living happily past age 90, after most friends of the same generation have died, is the continued willingness to forge friendships with younger people.

All of these qualities contributed to Ernst Mayr's scientific greatness and his productivity. They also lie at the root of the love felt for him by several generations of colleagues and friends.

Jared Diamond

Jared Diamond is in the Department of Geography, University of California, Los Angeles, California 90095-1524, USA.
e-mail: jdiamond@geog.ucla.edu

Climate change

Ancient floods in New York

Geology **33**, 89–92 (2005)

The cold spells that interrupted the last deglaciation around 13,000 years ago are thought to have been initiated by massive releases of meltwater from ice sheets into the North Atlantic Ocean, with consequent effects on oceanic heat circulation.

Jeffrey P. Donnelly *et al.* now claim to have mapped the course of one such event 13,350 years ago. It seems that meltwater from the Laurentide Ice Sheet in North America gathered in several glacial lakes around the Hudson Valley of the northeastern United States. Radiocarbon dates of sediments imply that one lake, Glacial Lake Iroquois, drained into Glacial Lakes Vermont and Albany as the ice sheet receded. This input overwhelmed a dam of glacial deposits at the lower end of Glacial Lake Albany, now the narrows on the Hudson River in New York City, triggering a catastrophic flood down the Hudson Valley. A second influx of meltwater to the North Atlantic apparently came from Glacial Lake Candona, draining along the St Lawrence Valley.

Donnelly *et al.* estimate that the change in salinity of the North Atlantic would have been sufficient to reduce the ocean's thermohaline circulation, and trigger climate cooling.

Philip Ball

Cell biology

Stress response

Science **307**, 935–939 (2005)

Endoplasmic reticulum (ER) stress is a feature of diabetes, Alzheimer's disease and viral infections. It is a consequence of cells' inability to fold proteins properly, ultimately resulting in cell death. After screening some 19,000 compounds, researchers have unearthed a molecule that protects rat cells from this untimely end.

The molecule, which Michael Boyce *et al.* have christened salubrinal, blocks the dephosphorylation of the α -subunit of eukaryotic translation initiation factor 2 (eIF2 α). Phosphorylation of eIF2 α is an early, protective part of the ER stress response. But it is usually only transient, and it seems that salubrinal helps to prevent the ensuing dephosphorylation.

Boyce *et al.* also show that salubrinal can slow the course of infection with herpes simplex virus. The virus's ability to make its own proteins is hampered if eIF2 α is in the phosphorylated state — hence the ability of salubrinal to impede viral replication.

Michael Hopkin

Anthropology

Climbing the social ladder

Proc. R. Soc. Lond. B doi:10.1098/rspb.2004.2970 (2005)

How many Christmas cards did you send last year — and how close are you to the recipients? W.-X. Zhou *et al.* have used their own data on this seasonal activity, along with other records, to study patterns in how people group together.

In different cultures all over the world, social clusters seem to fall into broad categories, always of roughly the same size: a core of 3–5 people who would be consulted at times of distress; a group of 12–20 individuals with quite close ties; bands of 30–50 people who for example work or (in traditional societies) camp together; and so on.

By analysing records gathered from many different countries and societies, both large and small, together with their Christmas-card data, Zhou *et al.* find that, no matter what the culture, each level of the social hierarchy seems to be about three times larger than the previous one.

Why such constraints might exist is puzzling. Perhaps group size is restricted by cognitive capacity (being able to remember



everyone's relationships with everyone else, for example), or perhaps it is restricted merely by the time it takes to keep in touch with other people.

Helen Dell

Astrophysics

Galaxies flare up

Astrophys. J. preprint at <http://arXiv.org/astro-ph/0501520> (2005)

The centres of active galaxies can generate large amounts of energy, up to a significant fraction of that of their host galaxy. These 'active galactic nuclei' (AGNs) are thought to be powered by matter falling into supermassive black holes.

Using the Subaru Telescope on the summit of Mauna Kea, Hawaii, Tomonori Totani *et al.* have found six very dim AGNs that emit flares of visible light on a timescale of a few days. AGNs generally vary over several months, although scientists have recently found that Sagittarius A*, the AGN at the centre of our own Galaxy, emits near-infrared flares that change from hour to hour.

Totani and colleagues' observations indicate that rapid flaring may be much more common than previously thought. The researchers believe that the AGNs are black holes around 100 million times heavier than the Sun. They suggest that the observed flares are coming from the event horizon, the very edge of the gravitational precipice that surrounds a black hole.

Rapid optical flaring activity may often be hidden in apparently normal galaxies, the authors say, and could be revealed by similar surveys that look for short-term variability in galaxies.

Mark Peplow

Developmental biology

Youth gone by

Cell **120**, 383–393 (2005)

In fruitflies, cells can be diverted from making legs and instead form wing tissue. This switch in fate is called transdetermination, and Anne Sustar and Gerold Schubiger provide new insights into how it works.

The authors analysed imaginal disc cells in fruitflies. These cells are usually 'fated' to become a specialized type in adults (leg or wing cells, for instance). However, when the fly is injured, cells near the wound are stimulated to divide and form a 'regeneration blastema'. If the blastema arises in a defined region of the disc, the cells can switch their fate. But do they do so directly or by first regressing to a developmentally less mature starting point?

To find out, Sustar and Schubiger determined the proportion of blastema cells in the various different stages of the cell cycle. They point out that those undergoing transdetermination adopt an entirely new type of cell-cycle profile, which involves getting bigger and spending longer in the DNA-replication phase. The duo offers this as evidence that these cells do not have to revert to their youth in order to take on a different fate. The findings may provide clues to the developmental potential of adult stem cells.

Roxanne Khamsi

Migratory bands give crickets protection

Predators are not so lucky in picking out a flightless insect from a marching mass.

Mormon crickets and juvenile locusts form huge migratory bands — millions of individuals march in unison across the landscape^{1–3} and devastate vast agricultural areas, but little is known about why these bands form. Here we use radiotelemetry to show that band membership benefits these insects by greatly reducing the probability that they will become victims of predators. It is likely that migratory banding has evolved because it gives substantial protection to individuals within the group.

Protection against predators through mechanisms such as enhanced early detection, predator confusion and dilution of risk (the 'selfish herd' effect) has been proposed to account for the evolution and maintenance of large aggregations in animals^{4,5}. The anti-predator benefits of migratory-band formation have previously been assumed to be negligible³, but have never been quantified — largely because of the difficulty of tracking individuals within bands and the rarity of witnessing predation events^{1,3,4}. We have used radiotelemetry to overcome these problems.

The mormon cricket, *Anabrus simplex*, is a flightless katydid⁶, native to western North



Figure 1 A migratory band of flightless mormon crickets (*Anabrus simplex*) crossing a dirt road in northeastern Utah, United States.

America, which forms spectacular migratory bands. These can be more than 16 kilometres long and several kilometres wide, with each square metre containing dozens of insects that walk up to 2 km a day^{2,7} (Fig. 1). In a replicated mark-recapture experiment, we compared the survival of individual mormon crickets in naturally occurring migratory bands with that of individuals transplanted from the band to nearby sites; mormon-cricket bands had previously travelled through these sites, but they were empty at the time of the experiment. Individuals were located using small radiotransmitters⁸ (Fig. 2a), which enabled us to establish their fate with accuracy (for methods, see supplementary information).

Insects that had been translocated from the band suffered 50–60% mortality due to predation over just two days in three replicate experiments (Fig. 2b). By contrast, we observed no mortality among individuals in the band during the same period, including among those crickets moved to control for any transportation effect (Life Table Survival analysis, Gehan's generalized Wilcoxon test: P was 0.0118, 0.0047 and 0.0026 for three replicates, respectively; Fig. 2b).

Predation was evident from recovered radiotransmitters that had been partially chewed (Fig. 2a, inset) and often had body parts still attached; they were retrieved from trees and burrows, suggesting that birds and rodents are likely predators². Two radiotransmitters were never recovered, presumably because of predator damage or removal from our detection range.

The evolution and maintenance of migratory-band formation requires the fitness benefits of group living to outweigh its costs⁴. Our results indicate that band formation confers a major anti-predator benefit, but what are the costs? Although difficult to quantify, intraspecific competition for resources is assumed to be a cost of group living and can account for

group movement to exploit new resources^{1,3,4}. Intraspecific competition in mormon-cricket bands mediates a food-stress-induced reversal of courtship roles at high population densities^{7,9}. Mormon crickets are also notoriously cannibalistic and likely to attack immobile conspecifics^{2,7}. Without discounting other hypotheses⁴, our results indicate that aggregation and constant movement protect band members from predators while reducing costs due to competition for resources and cannibalism. They also support a general anti-predator role for outbreaks such as those in periodical cicadas, where synchronized, mass emergence reduces predation on individuals by satiating local predators^{10,11}.

Gregory A. Sword*, Patrick D. Lorch†, Darryl T. Gwynne‡

*Northern Plains Agricultural Research Laboratory, US Department of Agriculture, Agricultural Research Service, Sidney, Montana 59270, USA
e-mail: gsword@sidney.ars.usda.gov

†Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3280, USA

‡Biology Department, University of Toronto at Mississauga, Mississauga, Ontario L5L 1C6, Canada

1. Uvarov, B. P. *Grasshoppers and Locusts* Vol. II (Centre for Overseas Pest Research, London, 1977).
2. MacVean, C. M. in *Integrated Pest Management on Rangeland: A Shortgrass Prairie Perspective* (ed. Capinera, J. L.) 116–136 (Westview, Boulder, Colorado, 1987).
3. Dingle, H. *Migration: The Biology of Life on the Move* (Oxford Univ. Press, Oxford, 1996).
4. Krause, J. & Ruxton, G. D. *Living in Groups* (Oxford Univ. Press, Oxford, 2002).
5. Hamilton, W. D. J. *Theor. Biol.* **7**, 1–52 (1971).
6. Sword, G. A. *Anim. Behav.* **69**, 437–444 (2005).
7. Gwynne, D. T. *Katydid and Bush-Crickets: Reproductive Behavior and Evolution of the Tettigoniidae* (Cornell Univ. Press, Ithaca, New York, 2001).
8. Lorch, P. D. & Gwynne, D. T. *Naturwissenschaften* **87**, 370–372 (2000).
9. Gwynne, D. T. *Ecology* **74**, 1406–1413 (1993).
10. Karban, R. *Ecology* **63**, 321–328 (1982).
11. Williams, K. S., Smith, K. G. & Stephen, F. M. *Ecology* **74**, 1143–1152 (1993).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

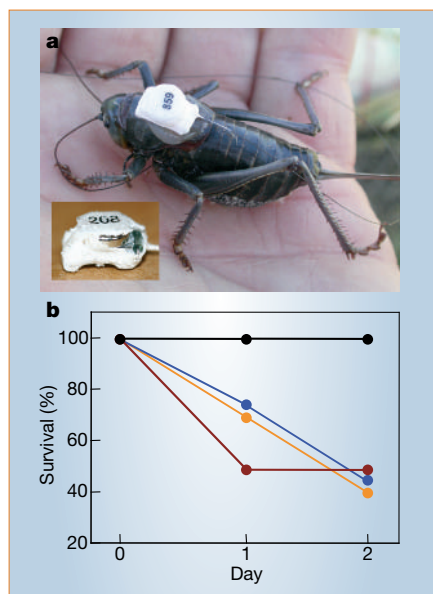


Figure 2 Radiotelemetric mark-recapture study reveals anti-predator benefits for insects in migratory bands. **a**, Female mormon cricket with a 0.45-g radiotransmitter glued to the pronotum. Inset, a recovered radiotransmitter showing evidence of predation, which in this case is likely to have resulted from chewing by a rodent. **b**, Survival of mormon crickets in a migratory band (black; $n = 10$ for all replicates) compared with that of conspecifics transplanted away from the band (three replicate experiments: red, $n = 10$; blue, $n = 20$; orange, $n = 20$). An equal number of males and females were used in each treatment and there was no effect of sex on survival (Cox regression, Wald statistic = 0.008, d.f. = 1, NS).

Chemical communication

Butterfly anti-aphrodisiac lures parasitic wasps

To locate their hosts, parasitic wasps can 'eavesdrop' on the intraspecific chemical communications of their insect hosts^{1–3}. Here we describe an example in which the information exploited by the parasitic wasp *Trichogramma brassicae* is a butterfly anti-aphrodisiac that is passed from male to female *Pieris brassicae* butterflies during mating, to render them less attractive to conspecific males^{4–6}. When the tiny wasp detects the odour of a mated female butterfly, it rides on her (Fig. 1) to her egg-laying sites and then parasitizes the freshly laid eggs. If this fascinating strategy is widespread in nature, it could severely constrain the evolution of sexual communication between hosts.

In two-choice olfactory bioassays (for methods, see supplementary information), females of the parasitic wasp *T. brassicae* showed a clear preference for odours of mated *P. brassicae* females or males (Fig. 2a). The wasps were more strongly attracted by the scents of male and mated female butterflies than of virgin females, but they did not discriminate between male and mated female butterflies. When offered separately against clean air, odours from male or mated female butterflies attracted the wasps significantly, but odours from virgin females did not (Fig. 2a).

In similar behavioural trials, we determined the concentration of the anti-aphrodisiac, benzyl cyanide⁵, that is attractive to the wasp. When given solvent-treated virgin females as an alternative, wasps were significantly attracted by odour from virgin female butterflies treated with 2 µg ($P=0.008$, Wilcoxon's matched pairs signed rank test) and 1 µg ($P=0.016$, Wilcoxon's matched pairs signed rank test) benzyl cyanide. This result indicates that *T. brassicae* is using the anti-aphrodisiac benzyl cyanide from *P. brassicae* as a foraging cue.

In further two-choice bioassays, we investigated whether wasps mount a mated female butterfly in response to the anti-aphrodisiac. When wasps were exposed to butterflies, they preferred to climb on to mated females rather



Figure 1 *Trichogramma brassicae* wasp (roughly 0.5 mm long), hitches a lift on a mated, female butterfly (*Pieris brassicae*).

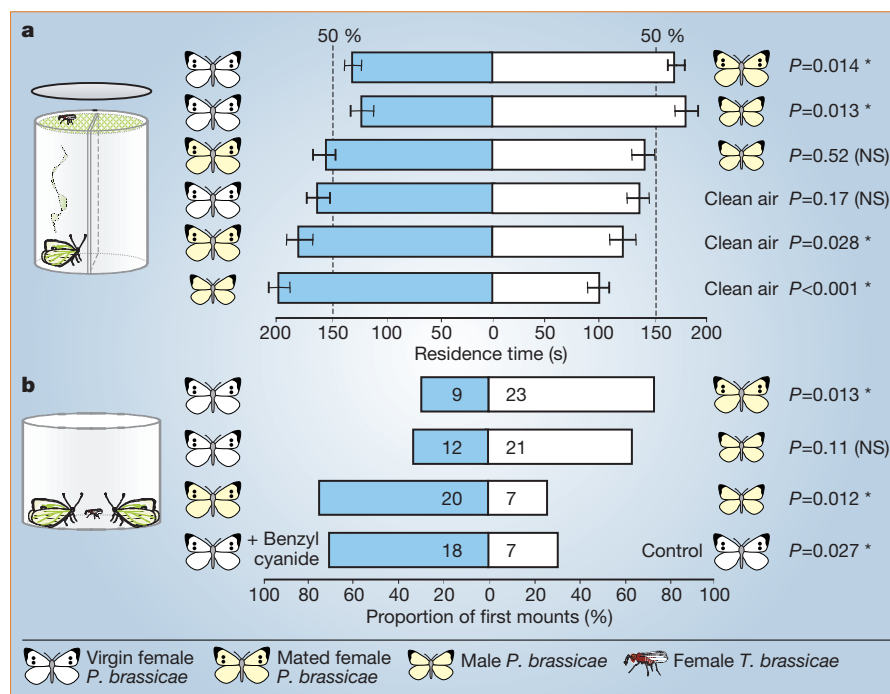


Figure 2 Response of *Trichogramma brassicae* parasitic wasps to signals of its host, the large cabbage-white butterfly *Pieris brassicae*. **a**, Mean residence time ($\pm$ s.e.m.) spent by wasps in two odour fields in a static olfactometer; $n=50$ wasps tested per combination (Wilcoxon's matched pairs signed ranks test). **b**, Proportion of first mounts (%) by *T. brassicae* wasps on adult butterflies. The number of responding wasps is shown inside each bar; $n=40$ wasps tested per combination (χ^2 test). *Trichogramma* size is disproportionately large as drawn (see Fig. 1). Asterisks indicate significant differences within the choice test; NS, not significant.

than on to males or virgin females (Fig. 2b). Discrimination between males and mated females in this assay was probably based on close-range cues other than benzyl cyanide. But when virgin female butterflies painted with 2 µg benzyl cyanide dissolved in hexane were offered against virgin female butterflies painted with hexane alone, wasps preferred to mount the butterflies treated with benzyl cyanide significantly more often than the controls (Fig. 2b).

We used a flight chamber to determine whether *T. brassicae* wasps are transported by the mated female butterfly to the host plant, the Brussels sprout (*Brassica oleracea* var. *gemmifera*), where they subsequently parasitize her freshly laid eggs. In total, 28 mated female butterflies were each seen carrying a single wasp; 14 wasps (50%) were found on the butterfly after it had landed on the host plant and had started laying eggs. Four of these wasps reached the host plant by descending from the butterfly, and two of them parasitized the new eggs after leaving the butterfly. Transportation (phoresy) on female butterflies therefore led to successful parasitism by *T. brassicae* in at least 7.1% of cases.

Phoresy is an ambush strategy^{7,8} that has been only rarely observed in *Trichogramma* spp. (ref. 9; and N. E. F. and M. E. H., unpublished observations). But host location by chemical espionage in combination with phoresy, which we describe here, may have evolved frequently in egg parasitoids. From an egg parasitoid's point of view, it is most adaptive in those species with limited flight,

a narrow time window for successful parasitism¹⁰, and/or a preference for hosts that lay egg clutches. From the host's perspective, egg parasitoids represent significant mortality factors¹¹ and so exert selection on butterfly intraspecific communication, seriously constraining its evolution. The link between genetic variation in this important trait and local parasitoid selection pressure warrants further investigation.

Nina E. Fatouros*†, Martinus E. Huigens†, Joop J. A. van Loont†, Marcel Dicke†, Monika Hilker*

*Institute of Biology, Freie Universität Berlin, 12163 Berlin, Germany

e-mail: hilker@zedat.fu-berlin.de

†Laboratory of Entomology, Wageningen University, P.O. Box 8031, 6700 EH Wageningen, The Netherlands

- Stowe, M. K., Turlings, T. C. J., Loughrin, J. H., Lewis, W. J. & Tumlinson, J. H. *Proc. Natl Acad. Sci. USA* **92**, 23–28 (1995).
- Arakaki, N., Wakamura, S. & Yasuda, T. *J. Chem. Ecol.* **22**, 1079–1085 (1996).
- Vinson, S. B. in *Insect Communication* (ed. Lewis, T.) 325–348 (Academic, London, 1984).
- Andersson, J., Borg-Karlson, A.-K. & Wiklund, C. *Proc. R. Soc. Lond. B* **267**, 1271–1275 (2000).
- Andersson, J., Borg-Karlson, A.-K. & Wiklund, C. *J. Chem. Ecol.* **29**, 1489–1499 (2003).
- Gilbert, L. E. *Science* **193**, 419–420 (1976).
- Clausen, C. P. *Annu. Rev. Entomol.* **21**, 343–368 (1976).
- Bruni, R., Sant'Ana, J., Aldrich, J. R. & Bin, F. J. *Insect Behav.* **12**, 165–173 (2000).
- Tabata, S. & Tamanuki, K. *Rev. Appl. Entomol. A* **29**, 95 (1940).
- Orr, D. B., Russin, J. S. & Boethel, D. J. *Can. Entomol.* **118**, 1063–1072 (1986).
- Hilker, M. & Meiners, T. (eds) *Chemecology of Insect Eggs and Egg Deposition* (Blackwell, Berlin, 2002).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Shape coexistence and triaxiality in the superheavy nuclei

S. Ćwiok^{1*}, P.-H. Heenen² & W. Nazarewicz^{3,4,5}

¹Institute of Physics, Warsaw University of Technology, ul. Koszykowa 75, PL-00662, Warsaw, Poland

²Service de Physique Nucléaire Théorique, Université Libre de Bruxelles, CP 229, B-1050 Brussels, Belgium

³Department of Physics and Astronomy, The University of Tennessee, Knoxville, Tennessee 37996, USA

⁴Physics Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37831, USA

⁵Institute of Theoretical Physics, Warsaw University, ul. Hoza 69, PL-00681, Warsaw, Poland

* Deceased

Superheavy nuclei represent the limit of nuclear mass and charge; they inhabit the remote corner of the nuclear landscape, whose extent is unknown. The discovery of new elements with atomic numbers $Z \geq 110$ has brought much excitement to the atomic and nuclear physics communities. The existence of such heavy nuclei hangs on a subtle balance between the attractive nuclear force and the disruptive Coulomb repulsion between protons that favours fission. Here we model the interplay between these forces using self-consistent energy density functional theory; our approach accounts for spontaneous breaking of spherical symmetry through the nuclear Jahn–Teller effect. We predict that the long-lived superheavy elements can exist in a variety of shapes, including spherical, axial and triaxial configurations. In some cases, we anticipate the existence of metastable states and shape isomers that can affect decay properties and hence nuclear half-lives.

The mere existence of superheavy elements has been a longstanding fundamental scientific problem. How can a nucleus with a large atomic number, such as $Z = 112$, survive the huge electrostatic repulsion between the protons? What are its physical and chemical properties? What is the extent of the superheavy region, that is, is there an upper limit on the number of neutrons and protons that can be bound into one cluster? Do there exist very-long-lived superheavy nuclei?

We know the answers to some of these questions¹. According to the nuclear shell model, a nucleon—that is, a proton or a neutron—moves in a common potential generated by all the other nucleons. Similar to an electron's motion in an atom, nucleonic orbits bunch together forming shells, and nuclei having filled nucleonic shells (nuclear 'noble gases') are exceptionally stable. This happens for specific 'magic' numbers of protons ($Z = 2, 8, 20, 28, 50$ and 82) and neutrons ($N = 2, 8, 20, 50, 82$ and 126). The quantum enhancement in nuclear binding due to the presence of nucleonic shells can be quantified in terms of the so-called shell energy². Although the magic nuclei have the largest shell energies, other nuclei can also be shell-stabilized, because the shell energy oscillates strongly with the number of nucleons.

By the end of the 1960s, it had been concluded that the existence of the heaviest nuclei with $Z > 104$ was primarily determined by the shell effects^{3–5}. These early calculations predicted the nucleus with $Z = 114$, $N = 184$ to be the centre of an island of long-lived superheavy nuclei (see, for example, the discussion in ref. 6). This result stayed practically unchallenged until the late 1990s, when more refined models, based on realistic effective nucleon–nucleon interactions, were applied to superheavy nuclei. The microscopic models are, however, still uncertain when extrapolating in Z and the mass number A . In particular, there is no consensus among theorists with regard to what should be the next doubly magic nucleus beyond ^{208}Pb ($Z = 82$, $N = 126$). In the superheavy nuclei the density of single-particle energy levels is fairly large, so small energy shifts, such as those due, for instance, to poorly known parts of nuclear interaction, can be crucial for determining the shell stability. This situation is similar to that encountered in the chemistry of superheavy elements, where the high density of single-electron states combined with relativistic effects make theoretical predictions

difficult⁷. Modern calculations suggest that the next magic proton shell should appear at higher proton numbers than previously thought: $Z = 120, 124$ or 126 , whereas for the neutrons, most calculations predict magic gaps at $N = 184$ or $N = 172$ (refs 6, 8, 9). In contrast to normal nuclei, however, the regions of enhanced shell effects in the superheavy region are generally expected to be fairly broad; that is, the magic gaps are fragile¹⁰.

At large values of Z and of the mass number $A = Z + N$, the electrostatic repulsion becomes so strong that the nuclear liquid drop becomes unstable to surface distortions⁴ and fissions. Because—as discussed below—the quantum shell energy may also favour shape deformation, many superheavy elements seem to be well deformed—that is, the deformed minimum is deep. Indeed, the measured α -decay energies, along with complementary syntheses of new neutron-rich isotopes of seaborgium ($Z = 106$) and hassium ($Z = 108$), have confirmed the special stability of the deformed nuclei with $N = 162$ predicted by theory^{11–13}.

Beautiful experimental confirmation of large quadrupole shape deformations in the heavy-element region comes from γ -ray spectroscopy around $Z = 102$, $N = 152$; namely, the identification of rotational bands in ^{254}No and its neighbours (see ref. 14). Figure 1 shows the deformation energies and quadrupole deformations for even–even heavy and superheavy nuclei calculated here. The largest ground-state shape elongations are indeed predicted at around ^{254}No . The well-deformed elongated (prolate) superheavy nuclei are separated from spherical elements with $Z = 114–126$, $N = 184$ by the region of weakly deformed, flattened (oblate) nuclei that are candidates for shape isomerism effects or triaxiality^{6,15,16}. Here we examined this transitional region, rich in structural phenomena.

Experimental status

The quest for heavy elements can be divided into several periods¹⁷. The first period (1896–1940) was characterized by the Curies' work, the first attempts to reach the elements beyond uranium by Fermi and Segre in 1934, and the discovery of nuclear fission in 1938. The Manhattan Project marked the second period (1940–1952), when plutonium became part of the periodic table. The third period (1955–1974) witnessed a Cold War competition between Russian and American laboratories to discover new elements.

The superheavy elements represent the fourth period¹³, which started with the first observations of elements $Z = 107$ – 109 at the Gesellschaft für Schwerionenforschung (GSI) in Germany in the beginning of the 1980s. Interest in those exotic species has been rekindled over the last decade, thanks to experimental progress in their synthesis^{18–21}. Isotopes of elements with the atomic number $Z = 110$ – 112 were discovered at GSI between 1994 and 1996¹⁹. Those elements are expected to be well-deformed (see Fig. 1) and their lifetimes have been found to be very short. For instance, the isotope ${}^A_Z = {}^{277}_{112}$ turned out to have a half-life of the order of 300 μ s. The element 110 (darmstadtium, Ds) was independently confirmed in 2002 by the Lawrence Berkeley Laboratory in the USA²⁰ and the Institute of Physical and Chemical Research (RIKEN) in Japan²¹. The confirmation of $Z = 111$ (roentgenium, Rg) and $Z = 112$ was also made at RIKEN. In a recent experiment, the Japanese physicists have reported the synthesis of $Z = 113$ ($A = 278$)²². The production cross-section was found to be rapidly decreasing with the atomic number (it is around 50 pb for $Z = 113$; 1 barn = 10^{-28} m²), so it was concluded that it would be very difficult to reach still heavier elements in ‘cold’-fusion reactions using lead or bismuth targets.

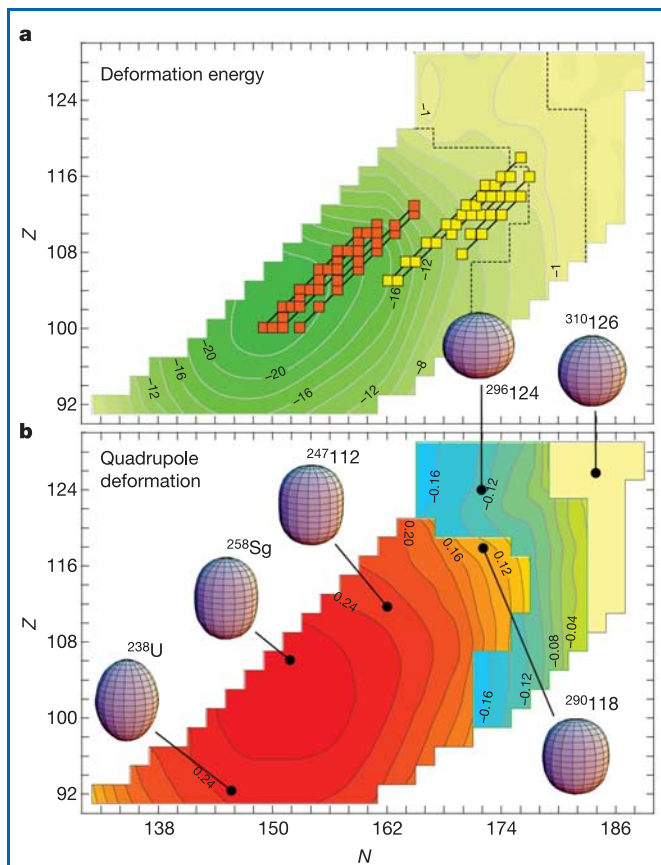


Figure 1 Deformation properties of even-even superheavy nuclei calculated self-consistently in the (N, Z) -plane with the SLy4 nuclear energy density functional. The centre of the shell stability is predicted around $N = 184$, $Z = 126$. **a**, Deformation energy (in MeV) defined as a difference between the ground-state energy and the energy at the spherical shape. The $Z = 110$ – 113 α -decay chains found at GSI and RIKEN are marked by red squares. The $Z = 118$, 116, 115 and 114 unconfirmed α -decay chains reported in Dubna are marked by yellow squares. **b**, Predicted ground-state mass quadrupole deformation β_2 (extracted using equation (7) of ref. 6 from the calculated axial mass quadrupole moment Q_{20}) and corresponding nuclear shapes for selected nuclei. Prolate shapes ($\beta_2 > 0$) are coloured red–orange, oblate shapes ($\beta_2 < 0$) are blue–green, and spherical shapes ($\beta_2 = 0$) are light yellow.

The use of ‘hot’-fusion evaporation reactions with the neutron-rich ${}^{48}\text{Ca}$ beam and actinide targets at the Joint Institute for Nuclear Research in Dubna (Russia) has resulted in measurements of several new elements with $Z = 113$ – 116 and 118 (refs 23, 24), including the isotopes of ${}^{286-290}_{114}$, ${}^{290-293}_{116}$, ${}^{294}_{118}$, and several new isotopes of $Z = 110$ and 112. According to Fig. 1, these nuclei, expected to have moderate deformations, belong to the region of shape transition from prolate- to oblate-deformed ground states. The most significant result is the observed increase of half-lives with increasing neutron number. For instance, when going from ${}^{282}_{112}$ ($N = 170$) to ${}^{285}_{112}$ ($N = 173$), the half-life increases from 1 ms to 34 s (ref. 25). This is consistent with the predicted increased stability of superheavy elements when approaching $N = 184$.

The hot-fusion measurements from Dubna still await confirmation by other laboratories. This is not going to be easy, because these newly synthesized nuclei form an isolated island that is not linked through α -decay chains with any known nuclei. Therefore, alternative techniques to identify the elements, such as chemistry or very precise mass measurements, are also being pursued.

Nuclear energy density functional

A theoretical framework aiming at the description of the structure of superheavy nuclei must fulfil several strict requirements. Most importantly, it must be general enough to be confidently applied to a region of the nuclear chart whose properties are largely unknown. Because the universal effective nucleon–nucleon interaction in heavy nuclei has not yet been derived microscopically, the preferred strategy is to use forces adjusted to selected experimental data. This must be done in a way general enough that the resulting effective

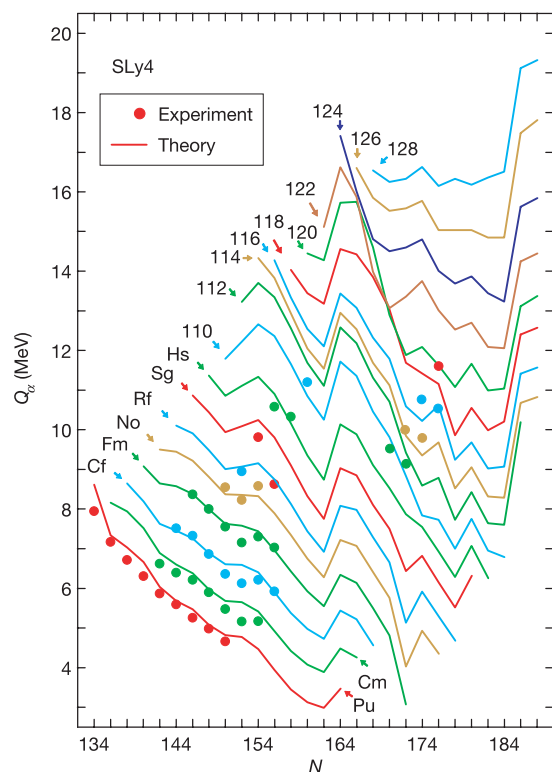


Figure 2 Q_α values for even-even nuclei with $96 \leq Z \leq 118$ obtained in the self-consistent calculations using the energy density functional SLy4. They are compared to experimental data (closed symbols), including the recent Dubna–Livermore data on the $Z = 116$ and 118 α -decay chains^{23,24}. The irregular behaviour of Q_α as a function of particle numbers can be attributed to shell effects and the resulting deformation changes. (After ref. 1.)

interaction can be used all over the nuclear chart. Another requirement for the method, of particular importance for the superheavy nuclei, is that it must treat correctly the unusually strong interplay between the nuclear and Coulomb forces. Finally, as we show below, studies of superheavy nuclei also require a method that can handle symmetry-breaking effects resulting in a large variety of intrinsic nuclear deformations.

These requirements are met by the density functional theory (DFT) in the formulation of Kohn and Sham²⁶. DFT is one of the most popular and successful quantum mechanical approaches to the many-body electronic structure calculations of molecular and condensed matter systems. For many years, it has also been used in many-body nuclear structure calculations²⁷. The main ingredient of the nuclear DFT is the energy density that depends on densities and currents representing distributions of nucleonic matter, spins, momentum and kinetic energy, as well as their derivatives (gradient terms). Such density functionals employed in self-consistent mean-field calculations are parameterized by means of about ten coupling constants that are adjusted to basic properties of nuclear matter and to selected data on finite nuclei. They are augmented by the pairing term, which describes nuclear superfluidity²⁸. When not corrected by correlation terms, standard nuclear energy functionals reproduce total binding energies with a root-mean-square (r.m.s.) error of the order of 2 to 4 MeV (ref. 29), and they usually perform quite well when applied to energy differences, radii and nuclear moments and deformations²⁷.

The self-consistency of the description is guaranteed by applying the Hartree–Fock–Bogoliubov method (HFB). Because all of the nucleons are treated on an equal footing and contribute to single-particle and pairing mean fields, a few hundred nucleonic wavefunctions have to be calculated which makes the computational effort quite demanding. In this work, we consider only ground

states of even–even isotopes. In this case, there is no intrinsic breaking of the time reversal symmetry. We shall use the standard nuclear energy density functional SLy4³⁰. Following ref. 6, in the pairing channel, we use the density-independent contact interaction and an approximate particle number projection is carried out by means of the Lipkin–Nogami method. The HFB equations are solved in coordinate space by discretization on a three-dimensional lattice³¹. Figure 2 illustrates the predictive power of our DFT calculations when applied to Q_α values (that is, energies of the α -particles emitted by radioactive heavy and superheavy nuclei). Our model reproduces the measured values quite accurately and hence we expect that the SLy4 functional can be reliably applied to as-yet-unobserved superheavy systems.

Deformed shapes of superheavy nuclei and shape coexistence

The mean-field description of nuclei is performed in a frame of reference attached to the nucleus, the intrinsic frame, in which the nucleus may acquire a deformed shape. The concept of shape deformation is ultimately related to the Jahn–Teller effect (spontaneous symmetry breaking) known in many areas of physics. Symmetry-breaking solutions may appear when, in a mean-field configuration respecting the original symmetries of the nuclear hamiltonian, degenerate single-particle orbits are strongly coupled to collective modes³². Figure 1 illustrates the importance of

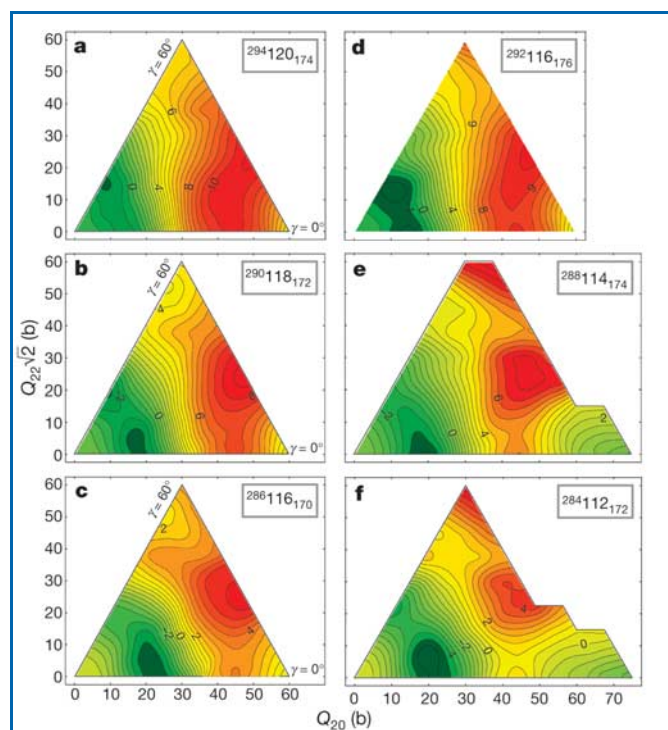


Figure 3 Potential energy surfaces of the members of the α -decay chains of $^{294}_{120}$ (a–c) and $^{292}_{116}$ (d–f) in the (Q_{20}, Q_{22}) plane calculated with the SLy4 energy density functional. It is seen that both α -decay sequences are associated with transition from oblate (or triaxial shapes) in the parent nuclei to prolate shapes in lighter daughter nuclei. The difference between contour lines is 0.5 MeV.

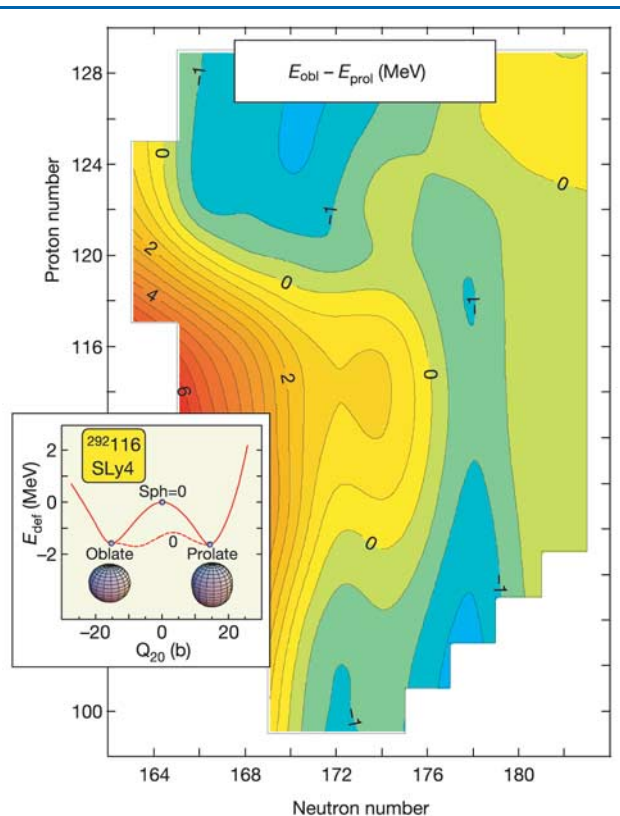


Figure 4 Contour map of the energy difference between oblate and prolate minima (or saddle points) in the energy surface of superheavy even–even nuclei. Positive (negative) numbers indicate prolate (oblate) ground states. The inset shows the total energy of the superheavy nucleus $^{292}_{116}$ (recently reported in ref. 23) as a function of the mass quadrupole moment Q_0 along the trajectory of axial shapes ($\gamma = 0^\circ$, solid line) and along the trajectory of triaxial shapes ($\gamma \neq 0^\circ$, dashed line). We can see that $^{292}_{116}$ is γ -soft, that is, the inclusion of triaxiality almost completely wipes out the barrier between prolate and oblate minima. However, additional separation between prolate and oblate configurations may exist owing to the presence of intruder orbitals (see text).

the symmetry-breaking effect in the superheavy region. Figure 1a displays the difference between the energies of the spherical and the deformed configurations for the even–even nuclei above uranium. For all the nuclei with at most 118 protons and 172 neutrons, the energy gain due to deformation is larger than 10 MeV. Compared to the total binding energy, the deformation energy seems to be small, but it determines the stability of the nucleus to two main decay modes in this region: α -emission and spontaneous fission.

Figure 1b shows the variation of the dimensionless mass quadrupole deformation β_2 with Z and N . All of the isotopes detected at GSI and RIKEN using cold-fusion reactions (marked by red squares in Fig. 1a) are located in a region of large prolate deformations. The heaviest isotopes reported in Dubna's hot-fusion experiments (yellow squares in Fig. 1a) are predicted to have smaller quadrupole moments and are either oblate or prolate, whereas their decay products lie in a region of rapidly changing shapes. As discussed below, nuclei from this region can be prone to triaxial distortions. (The importance of triaxial shapes in the heaviest and superheavy nuclei, especially in the context of fission, was noted in the early macroscopic–microscopic studies^{33–35}.) Only when we approach $N = 184$ do we expect superheavy nuclei that are spherical in their ground states.

That the shape of a nucleus depends crucially on its proton and neutron numbers is illustrated in Fig. 3, where the variation of the energy as a function of quadrupole moments Q_{20} and Q_{22} is plotted for three isotopes belonging to the α decay chains of $^{294}120$ (Fig. 3a–c) and $^{292}116$ (Fig. 3d–f). The quadrupole moments can be related to the Hill–Wheeler polar deformation parameters Q_0 and γ through the usual relations: $Q_{20} = Q_0 \cos \gamma$ and $Q_{22} = \frac{Q_0}{\sqrt{2}} \sin \gamma$. For weakly deformed shapes, Q_0 is simply proportional to the quadrupole deformation β_2 . The deformation parameter γ measures the degree of triaxiality. The $\gamma = 0^\circ$ limit corresponds to axial prolate shapes ($\beta_2 > 0$) while oblate axial shapes ($\beta_2 < 0$) appear at $\gamma = 60^\circ$. Intermediate values of γ are associated with triaxial shapes. The parent nuclei of both α -chains have near-oblate ground states; their daughters are calculated to be prolate, but with a very small barrier with respect to triaxial deformation γ . The ground states of the

lightest isotopes shown in Fig. 3 are expected to have near-prolate deformations.

To illustrate the prolate–oblate energy balance in the superheavy nuclei, Fig. 4 displays the energy difference between the oblate and prolate minima (which are sometimes only saddle points in the γ direction). The prolate ground states usually correspond to minima that are fairly deep; however, for the oblate ground states, the prolate configurations are, in most cases, excited by less than 1 MeV. Experimentally, one can thus expect the coexistence of $J^\pi = 0^+$ states close in energy that result from the mixing of configurations having different shapes. This phenomenon, called shape coexistence³⁶, is known in nuclei from several regions of the nuclear chart.

The single-neutron and single-proton energy levels plotted in Fig. 5 explain the origin of the shape change along the α -decay chains shown in Fig. 3 and the oblate–prolate competition of Fig. 4. The single-particle levels have been calculated for the nucleus $^{292}116$, but they are very similar for the six nuclei in question. The large gaps in the single-particle spectrum can be associated with increased stability of a given nuclear shape. For instance, the increased shell stability at particle numbers $N = 184$ and $Z = 126$ can be attributed to the corresponding spherical gaps in the single-particle spectrum. These gaps of about 2–3 MeV are, however, much smaller than the magic gaps in the known doubly magic nuclei^{9,10}. Therefore, deformation effects can play a particularly important role in the superheavy systems. It is the interplay between the pronounced oblate gaps at $Z = 120$ and $N = 172, 178$, and the large prolate gaps at $Z = 114, 116$ and $N = 174, 176$ that determines the microscopical shape coexistence and shape-transitional behaviour in this region.

Figure 5b and d display the single-particle energies as a function of γ for $Q_0 = 15$ b, which corresponds to the path connecting the prolate and the oblate minima. For $N \approx 176$ and $Z \approx 120$, the density of states around the Fermi level is very low in the γ direction, which explains the large softness of the calculated binding energy against γ (see inset in Fig. 4). Therefore, we expect that dynamical correlation effects associated with the large-amplitude collective motion in the γ direction will be large for most transitional nuclei

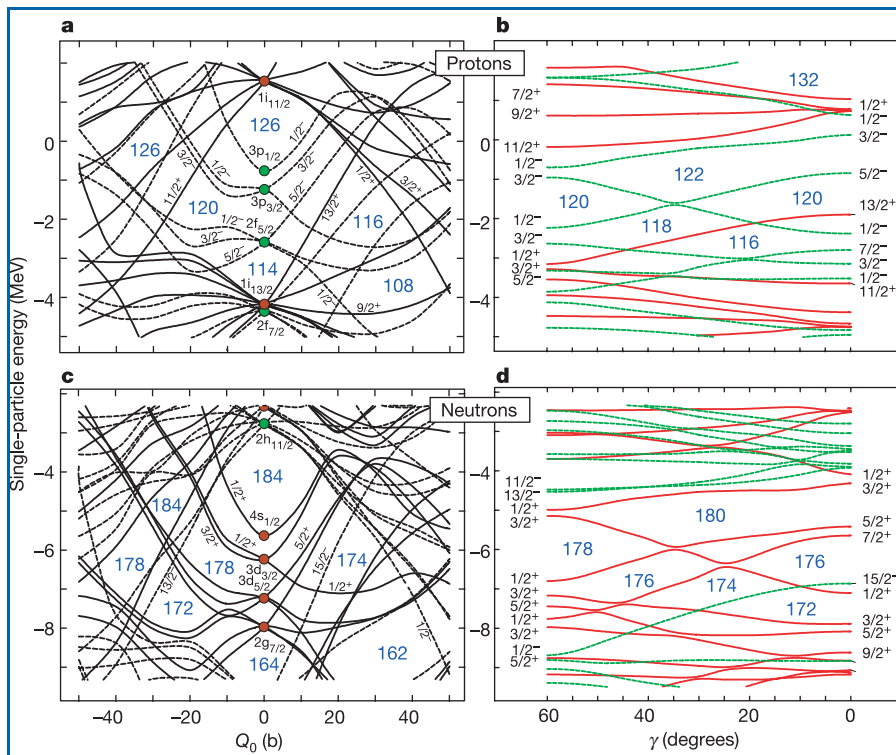


Figure 5 Single-proton (a, b) and single-neutron (c, d) energy levels in the superheavy nucleus $^{292}116_{176}$ obtained in self-consistent calculations with the SLy4 energy density functional. Solid and dashed lines mark positive-parity and negative-parity levels, respectively. **a, c**, Single-particle energies as a function of the mass quadrupole moment Q_0 ($\gamma = 0^\circ$). Positive (negative) values of Q_0 correspond to prolate (oblate) shapes. At the spherical point, nucleonic shells are labelled using the spherical quantum numbers (n, l), whereas the deformed orbitals are labelled using two quantum numbers Ω^π , where Ω is the projection of the single-particle angular momentum on the symmetry axis and π is parity. **b, d**, The dependence of single-particle energies on triaxial deformation γ (at a fixed value of $Q_0 = 15$ b). The Ω^π labels are shown only at $\gamma = 0^\circ$ (axial prolate shape) and $\gamma = 60^\circ$ (axial oblate shape). The large spherical and deformed subshell closures are indicated.

considered; they will lead to the spreading of the nuclear wavefunction in the direction of γ .

It has to be noted, however, that the softness of the total energy surface is not the whole story. There are numerous situations in which some specific single-particle structure can give rise to a strong separation between coexisting minima in the many-body configuration space, even if the potential barrier between the minima is very small or even nonexistent. Usually, the separation can be associated with an occupation of specific single-particle orbitals with very distinct quantum characteristics, the so-called 'intruder states'³⁶. In Fig. 5, we notice the presence of two such intruder levels, which at $\gamma = 0^\circ$ carry a very large projection of the single-particle angular momentum Ω on the symmetry axis. Those are the $\Omega^\pi = 13/2^+$ proton level (appearing just below the prolate gap at $Z = 120$) and the $\Omega^\pi = 15/2^-$ neutron level (appearing just below the prolate gap at $N = 176$). Both levels originate from the high- j unique parity spherical shells $\pi i_{13/2}$ and $\nu j_{15/2}$, respectively. For proton numbers $Z = 116$ and 118 , the $\Omega^\pi = 13/2^+$ proton level is clearly occupied at an oblate shape, and empty at a prolate shape. Likewise, at $N = 168$ – 172 the $\Omega^\pi = 15/2^-$ neutron is occupied (empty) at an oblate (prolate) configuration. In nuclei with those particle numbers, we expect coexistence between prolate and oblate shapes and that quantum correlations will not destroy the simple mean-field picture of shape coexistence. In other cases, our calculations suggest γ -softness and strong configuration mixing.

Discussion

The new theoretical results presented in this work were obtained in the self-consistent formalism based on a modern nuclear energy density functional. We performed calculations for the even–even heavy and superheavy nuclei with $Z \leq 128$ and $N \leq 188$. Our main conclusions can be summarized as follows:

- (1) The large electrostatic repulsion and large single-particle level density around the Fermi level mean that deformation effects play a particularly important role in the description of superheavy elements. Although the actinide and transfermium nuclei are known to possess well-deformed elongated shapes, the superheavy elements around $Z = 116$ and $N = 176$ are expected to exhibit coexistence of oblate and prolate shapes.
- (2) The inclusion of triaxiality can dramatically reduce the barrier separating prolate and oblate minima, leading to structures that are soft or unstable to triaxial distortions.
- (3) In most cases, the underlying single-particle configurations change gradually (adiabatically) along the triaxial energy surface. Hence, strong anharmonicities and large dynamical effects are likely. In some cases, however, (diabatic) prolate and oblate states are expected to be well separated in the configuration space, thanks to the presence of high- Ω intruder levels originating from high- j shells.
- (4) The heaviest isotopes recently reported in Dubna are predicted to belong to a transitional region strongly influenced by dramatic shape changes and/or triaxial softness. Such shape effects are expected to affect α -decay lifetimes in two ways. First, the deformation affects the energy of α -decay, Q_α . Second, we expect transitions involving parent and daughter nuclei with different shapes to be hindered. This may explain the observation of longer half-lives for some nuclei within these decay chains. Consequently, the existence of shape isomers in the superheavy nuclei may make the identification of the new species more difficult. $\square$

doi:10.1038/nature03336.

1. Heenen, P.-H. & Nazarewicz, W. Quest for superheavy nuclei. *Europhys. News* **33**, 1–9 (2002).

2. Strutinsky, V. M. Shells in deformed nuclei. *Nucl. Phys. A* **122**, 1–33 (1967).
3. Sobczewski, A., Gareev, F. A. & Kalinkin, B. N. Closed shells for $Z > 82$ and $N > 126$ in a diffuse potential well. *Phys. Lett. B* **22**, 500–502 (1966).
4. Myers, W. D. & Swiatecki, W. J. Nuclear masses and deformations. *Nucl. Phys.* **81**, 1–60 (1966).
5. Nilsson, S. G. *et al.* On the nuclear structure and stability of heavy and superheavy elements. *Nucl. Phys. A* **131**, 1–66 (1969).
6. Ćwiok, S., Dobaczewski, J., Heenen, P.-H., Magierski, P. & Nazarewicz, W. Shell structure of the superheavy elements. *Nucl. Phys. A* **611**, 211–246 (1996).
7. Schwerdtfeger, P. & Seth, M. B. in *Encyclopedia of Computational Chemistry* Vol. 4, 2480–2494 (Wiley, New York, 1998).
8. Bender, M., Rutz, K., Reinhard, P.-G., Maruhn, J. A. & Greiner, W. Shell structure of superheavy nuclei in self-consistent mean-field models. *Phys. Rev. C* **60**, 034304 (1999).
9. Kruppa, A. T. *et al.* Shell corrections of superheavy nuclei in self-consistent calculations. *Phys. Rev. C* **61**, 034313 (2000).
10. Bender, M., Nazarewicz, W. & Reinhard, P.-G. Shell stabilization of super- and hyperheavy nuclei without magic gaps. *Phys. Lett. B* **515**, 42–48 (2001).
11. Ćwiok, S., Pashkevich, V. V., Dudek, J. & Nazarewicz, W. Fission barriers of transfermium elements. *Nucl. Phys. A* **410**, 254–270 (1983).
12. Möller, P. & Nix, J. R. Stability of heavy and superheavy elements. *J. Phys. G* **20**, 1681–1747 (1994).
13. Armbruster, P. On the production of superheavy elements. *Acta Phys. Polon. B* **34**, 1825–1866 (2003).
14. Julin, R. In-beam spectroscopy of heavy actinides. *Nucl. Phys. A* **685**, 221–232 (2001).
15. Ren, Z. Shape coexistence in even-even superheavy nuclei. *Phys. Rev. C* **65**, 051304 (2002).
16. Buryanich, T., Bender, M., Maruhn, J. A. & Reinhard, P.-G. Systematics of fission barriers in superheavy elements. *Phys. Rev. C* **69**, 014307 (2004).
17. Hoffman, D. C., Ghiorso, A. & Seaborg, G. T. *The Transuranium People* (World Scientific, River Edge, New Jersey, 2000).
18. Hofmann, S. & Münzenberg, G. The discovery of the heaviest elements. *Rev. Mod. Phys.* **72**, 733–767 (2000).
19. Hofmann, S. *et al.* Properties of heavy nuclei measured at the GSI SHIP. *Nucl. Phys. A* **734**, 93–100 (2004).
20. Ginter, T. N. *et al.* Confirmation of production of element 110 by the $^{208}\text{Pb}(^{64}\text{Ni},n)$ reaction. *Phys. Rev. C* **67**, 064609 (2003); erratum **C68**, 029901 (2003).
21. Morita, K. *et al.* Status of heavy element research using GARIS at RIKEN. *Nucl. Phys. A* **734**, 101–108 (2004).
22. Morita, K. *et al.* Experiment on the synthesis of element 113 in the reaction $^{209}\text{Bi}(^{70}\text{Zn},n)^{278}113$. *J. Phys. Soc. Jpn.* **73**, 2593–2596 (2004).
23. Oganessian, Yu. Ts. *et al.* Heavy element research at Dubna. *Nucl. Phys.* **734**, 109–123 (2004).
24. Moody, K. & the Dubna–Livermore Collaboration. Superheavy element isotopes, decay properties. *Nucl. Phys. A* **734**, 188–191 (2004).
25. Oganessian, Yu. Ts. *et al.* Measurements of cross sections for the fusion-evaporation reactions $^{244}\text{Pu}(^{48}\text{Ca},xn)^{292-x}114$ and $^{245}\text{Cm}(^{48}\text{Ca},xn)^{293-x}116$. *Phys. Rev. C* **69**, 054607 (2004).
26. Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A* **140**, 1133–1138 (1965).
27. Bender, M., Heenen, P.-H. & Reinhard, P.-G. Self-consistent mean-field models for nuclear structure. *Rev. Mod. Phys.* **75**, 121–180 (2003).
28. Perlínska, E., Rohozinski, S. G., Dobaczewski, J. & Nazarewicz, W. Local density approximation for proton-neutron pairing correlations: formalism. *Phys. Rev. C* **69**, 014316 (2004).
29. Stoitsov, M. V., Dobaczewski, J., Nazarewicz, W., Pittel, S. & Dean, D. J. Systematic study of deformed nuclei at the drip lines and beyond. *Phys. Rev. C* **68**, 054312 (2003).
30. Chabanat, E., Bonche, P., Haensel, P., Meyer, J. & Schaeffer, R. A Skyrme parametrization from subnuclear to neutron star densities. Part II: nuclei far from stabilities. *Nucl. Phys. A* **635**, 231–256 (1998).
31. Bonche, P., Flocard, H., Heenen, P.-H., Krieger, S. J. & Weiss, M. S. Self-consistent calculation of triaxial deformations: application to the isotopes of Kr, Sr, Zr, and Mo. *Nucl. Phys. A* **443**, 39–63 (1985).
32. Nazarewicz, W. Microscopic origin of nuclear deformations. *Nucl. Phys. A* **574**, 27c–49c (1994).
33. Pashkevich, V. V. The energy of non-axial deformation in heavy nuclei. *Nucl. Phys. A* **133**, 400–404 (1969).
34. Larsson, S. E., Ragnarsson, I. & Nilsson, S. G. Fission barriers and the inclusion of axial asymmetry. *Phys. Lett. B* **38**, 269–273 (1972).
35. Götz, U., Pauli, H. C. & Junker, K. Influence of asymmetric distortions on fission barriers. *Phys. Lett. B* **39**, 436–438 (1972).
36. Wood, J. L., Heyde, K., Nazarewicz, W., Huysse, M. & Van Duppen, P. Coexistence in even-mass nuclei. *Phys. Rep. C* **215**, 101–201 (1992).

Acknowledgements We thank M. Bender, J. Dobaczewski and M. Stoyer for discussions. This work was supported in part by the US Department of Energy, by the National Nuclear Security Administration under the Stewardship Science Academic Alliances programme, by the Belgian Science Policy Office, by NATO, and by the Polish Committee for Scientific Research (KBN).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.-H.H. (phheenen@ulb.ac.be).

How sleep affects the developmental learning of bird song

Sébastien Derégnaucourt¹, Partha P. Mitra², Olga Fehér¹, Carolyn Pytte³ & Ofer Tchernichovski¹

¹Department of Biology, City College, City University of New York, New York 10031, USA

²Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

³Department of Biology, Wesleyan University, Middletown, Connecticut 06459, USA

Sleep affects learning and development in humans and other animals, but the role of sleep in developmental learning has never been examined. Here we show the effects of night-sleep on song development in the zebra finch by recording and analysing the entire song ontogeny. During periods of rapid learning we observed a pronounced deterioration in song structure after night-sleep. The song regained structure after intense morning singing. Daily improvement in similarity to the tutored song occurred during the late phase of this morning recovery; little further improvement occurred thereafter. Furthermore, birds that showed stronger post-sleep deterioration during development achieved a better final imitation. The effect diminished with age. Our experiments showed that these oscillations were not a result of sleep inertia or lack of practice, indicating the possible involvement of an active process, perhaps neural song-replay during sleep. We suggest that these oscillations correspond to competing demands of plasticity and consolidation during learning, creating repeated opportunities to reshape previously learned motor skills.

In humans and other animals, sleep facilitates learning and the consolidation of memory^{1–11}. Experimental studies have shown reactivation of neuronal populations during post-training sleep^{2,11}. For example, in a recent human functional magnetic resonance imaging study, hippocampal areas that were activated during spatial learning were reactivated during sleep, and the amount of this reactivation predicted the improvement in performance during the next day¹¹. These and other studies have emphasized the role of sleep in adults, whereas the possible effect of sleep in developmental learning has not been investigated, even though sleep can affect synaptic plasticity and brain development¹². Furthermore, children with persistent sleep problems are more likely to have behavioural problems¹³.

Developmental learning (for example speech acquisition in human infants¹⁴) takes place during early life and has effects that last the entire lifetime of the individual. Developmental learning is difficult to study because behavioural changes span many time-scales. The challenge is to relate behavioural changes that may occur within hours as well as across daily cycles of wakefulness and sleep to the outcome observed in the adult individual. The study of developmental song learning in birds provides a unique model system for examining this process in detail.

Like humans, songbirds have innate predispositions to imitate complex vocalizations^{14,15}. Zebra finch males develop their song between 30 and 90 days after hatching, a time known as the sensitive period for vocal learning¹⁶. At about day 30 after hatching, young males start producing unstructured sounds (subsong). The onset of vocal learning (after exposure to a song model) is marked by a rapid emergence of structured sounds^{17,18}. To learn a song, the bird has to compare these sounds with a memory template of the song model using auditory feedback¹⁹. Learning is achieved by transforming and differentiating prototype sounds until they resemble the different syllables of the song model^{17,18}. Vocal changes occur at both syllabic and syntactical levels^{17,18} in conjunction with anatomical and functional changes in the brain^{20,21}. Functional changes that are relevant to song learning are driven by changes in behavioural state. For example, neuronal auditory responses to song playbacks are modulated by wakefulness and sleep^{22–25}. Furthermore, some neurons of the premotor song nuclei show spontaneous bursting during sleep, similar to the pattern of activity in the awake, singing bird^{26,27}. This suggests the possibility of song rehearsal during

sleep²⁸. However, until now there has been no direct evidence that sleep can affect vocal learning. In this study we recorded the entire song development and examined in detail how vocal learning progresses through cycles of wakefulness and sleep and attempted to relate the transitory effect of sleep episodes to the long-term outcome of developmental learning.

Quantification of vocal changes and vocal learning

To allow continuous measurements we recorded the entire song development and segmented the developing songs into syllable units (about 1 million syllables per bird). We summarized the structure of each syllable by an array of acoustic features. Starting from continuous measurements of features²⁹, we computed means and variances of each feature across the syllable. These means and variances are the ‘syllable features’ in our subsequent analysis: duration, mean pitch, mean frequency modulation (FM), mean Wiener entropy (the entropy being a measure of the diversity of power across frequency), mean goodness of pitch, mean continuity (continuity being a measure of spectro-temporal continuity of frequency traces), variance of pitch, variance of entropy, variance of FM, variance of goodness of pitch and variance of continuity. Shortly after the onset of training with song playbacks¹⁷, we observed a rapid emergence of syllable types³⁰ (clusters; Fig. 1). The development of each cluster was traced semi-automatically³¹. Song developmental learning was then assessed by two different methods. First, we examined changes in means and variances of feature values collapsed across syllables. Vocal changes were measured within each cluster by comparing a sample of 100 syllables produced one day with another sample produced at a different time. We estimated the difference between two samples of songs by averaging vocal changes across all features of all clusters produced by the bird. Second, to estimate vocal learning, we measured similarity to the song model played to the bird, based on a more detailed sliding comparison of feature values in which intra-syllabic temporal relationship between feature values is preserved²⁹.

Vocal changes during night-sleep

We first trained 12 male zebra finches to imitate a song model from day 43 after hatching. To establish a baseline, we compared differences (absolute percentage of change) across two random samples of 100 songs produced during the same day (Fig. 2a, b,

green curve and bars). The difference between those samples is an estimate of the measurement error. We then examined vocal changes from one day to the next, comparing a random sample of 100 songs produced during one day with a random sample produced during the next day. The day-to-day vocal changes were higher than baseline (Wilcoxon, $P < 0.01$) and decreased during song development.

We then examined vocal changes that occurred during night-sleep by comparing the last 100 songs produced during the evening of one day with the first 100 songs produced in the morning of the next day (red curve). Strikingly, the evening-to-next-morning changes were higher than the day-to-day changes ($P < 0.01$). Similar results were obtained in a bird trained by a live tutor (Fig. 2b, striped bars). In adult birds (1 year old) we still observed that the day-to-day variations were larger than the within-day baseline, confirming retention of some plasticity^{32,33}. However, transitory changes in the morning were not observed (see below).

Because vocal changes after 12 h of night-sleep were larger than the overall changes that occurred during 24 h, vocal changes must be non-monotonic; that is, they must oscillate during a daily cycle. Indeed, we observed strong daily oscillations in syllable features, most pronouncedly in variance features that capture the richness of acoustic structure within a syllable (Supplementary Table 1). Figure 3 shows how values of Wiener entropy variance (EV) increased from day to day during development (Fig. 3a, b). Strong daily oscillations were observed shortly after training started (Fig. 3c, d), and decreased thereafter (Fig. 3e, f). Note that the decrease in EV after night-sleep was opposite to the overall trend of

increase. Pooling all the variance features showed that during early syllable development (Fig. 3h, red curve) the diversity in syllable features was low in the morning, indicating a decrease in intra-syllabic structure after the night (repeated-measures analysis of variance (ANOVA), $P < 0.001$). Syllable structure recovered with 2–3 h of intense morning singing. Suspending training with song playbacks in the morning did not affect the recovery of song structure ($n = 5$, Wilcoxon, not significant (n.s.)). Daily oscillations in song structure were much weaker during late syllable development (blue curve), and undetectable in 1-year-old birds (green curve). Birds trained when adult, starting from day 90 ($n = 8$), exhibited substantial vocal changes but showed very weak post-sleep changes. We compared post-sleep EV changes across age groups, selecting syllables that showed similar magnitudes of change ($\pm$ s.e.m.) from one day to the next ($5.1 \pm 0.3\%$ versus $5.3 \pm 0.5\%$ EV increase per day in birds trained from day 43 and in birds trained from day 90, respectively). We found strong oscillations, with a post-sleep decrease in EV ($-16.7 \pm 1.5\%$) in birds trained from day 43, and a very weak decrease ($-2.4 \pm 2\%$) in birds trained from day 90 (Mann–Whitney, $P < 0.01$; Supplementary Fig. 1 and Supplementary Data).

The decrease in syllable structure in the morning may suggest that vocal changes during the day progress with the overall developmental trend, whereas vocal changes after night-sleep oppose the developmental trend. To test this hypothesis we assigned a sign to vocal changes during night-sleep, by reference to the overall developmental trend: positive if in the same direction, negative if opposite. For example, if syllable mean pitch decreased during

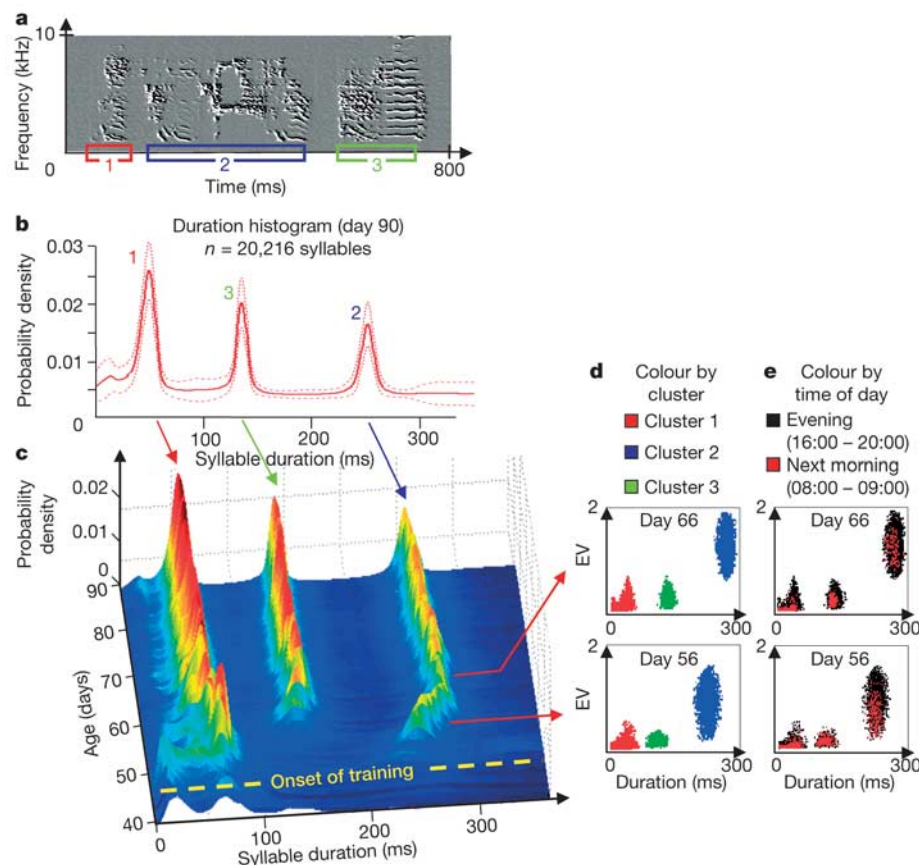


Figure 1 Tracing vocal changes. **a**, Spectral derivatives²⁹ of adult song motif with three syllables. **b**, Smoothed histogram of syllable durations⁴⁶ in an adult bird ($\pm 95\%$ confidence interval). Each peak corresponds to a song-syllable type of distinct duration. **c**, To trace the development of syllable types we plot developmental maps in which each row represents the duration histogram of syllables during one day. Syllable types (ridges)

emerged shortly after training. **d**, Plotting two-dimensional distribution (duration versus Wiener EV) shows syllable types as clusters (unclustered syllables are not shown). Changes in the position of clusters reveal vocal changes. **e**, Colour-coding circadian time shows vocal changes after night-sleep.

night-sleep but increased during development, we assigned change during night-sleep as negative. Note that if vocal changes during night-sleep were due to random perturbations, the net change across many samples should approach zero. As shown in Fig. 2c, analysis across the 12 birds (including all features of all syllables) confirmed that the net effect of vocal changes after night-sleep was negative. Furthermore, deterioration was observed in both mean and variance feature values (Supplementary Table 1) and was maximal from day 50 to day 55. To separate the effect of training from that of chronological age, we repeated the experiment in birds

trained from day 60 (Fig. 2d, $n = 6$). As shown, delaying the onset of training delayed the effect (see Supplementary Data and Supplementary Fig. 2). Isolated birds ($n = 6$) that were kept in the same conditions but were not trained did not show a significant trend (Fig. 2e, f). The magnitude of post-sleep deterioration differed significantly across the groups: highest when training started early, lower when training started later, and lowest in the untrained birds (Fig. 2f, median test, $P < 0.05$). Overall, song was less structured and more primitive after night-sleep but only during days of rapid learning.

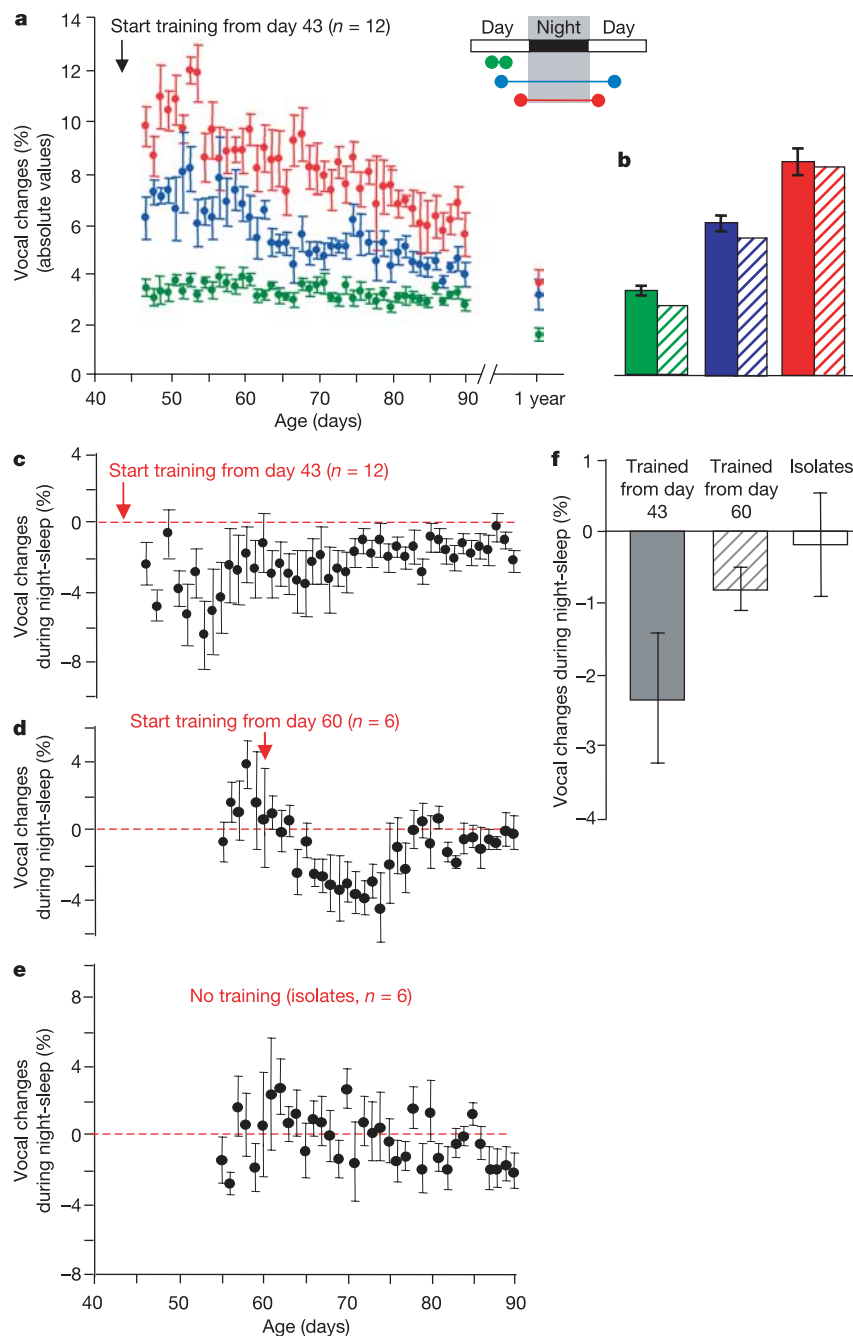


Figure 2 Vocal changes during night-sleep. **a**, Vocal changes (absolute values; medians $\pm$ s.e.m.) observed during song development in 12 birds trained from day 43. Green, vocal changes across random samples obtained during the same day (baseline); blue, vocal changes from one day to the next; red, vocal changes from one evening to the next morning (during night-sleep). **b**, The same data, with medians across song development (full bars) and in a live tutored bird (striped bars). **c–e**, Vocal changes during

night-sleep with reference to the overall developmental trend, in birds trained from day 43 (**c**), in birds trained from day 60 (**d**) and in untrained (isolated) birds (**e**). **f**, The same data as in **c–e**, showing medians across song development (median test, $P < 0.05$ across groups and post hoc between birds trained from day 43 and isolates; results are medians; error bars are s.e.m.).

Post-sleep deterioration and learning

We now turn to measurements of vocal learning and examine the eventual similarity to the model song (measured on day 90) against the magnitude of night-sleep deterioration. Similarity analysis is based on comparing detailed temporal structure across sliding windows of two sounds, yielding similarities as a percentage

measure²⁹. The overall magnitude of post-sleep deterioration during development is positively correlated with the eventual similarity to the model song (Fig. 4a, Spearman $\rho = 0.6$, $P = 0.038$; note, however, that correlation does not prove causality—see further analysis in Supplementary Data). In contrast, the net change in feature values during the entire learning period did

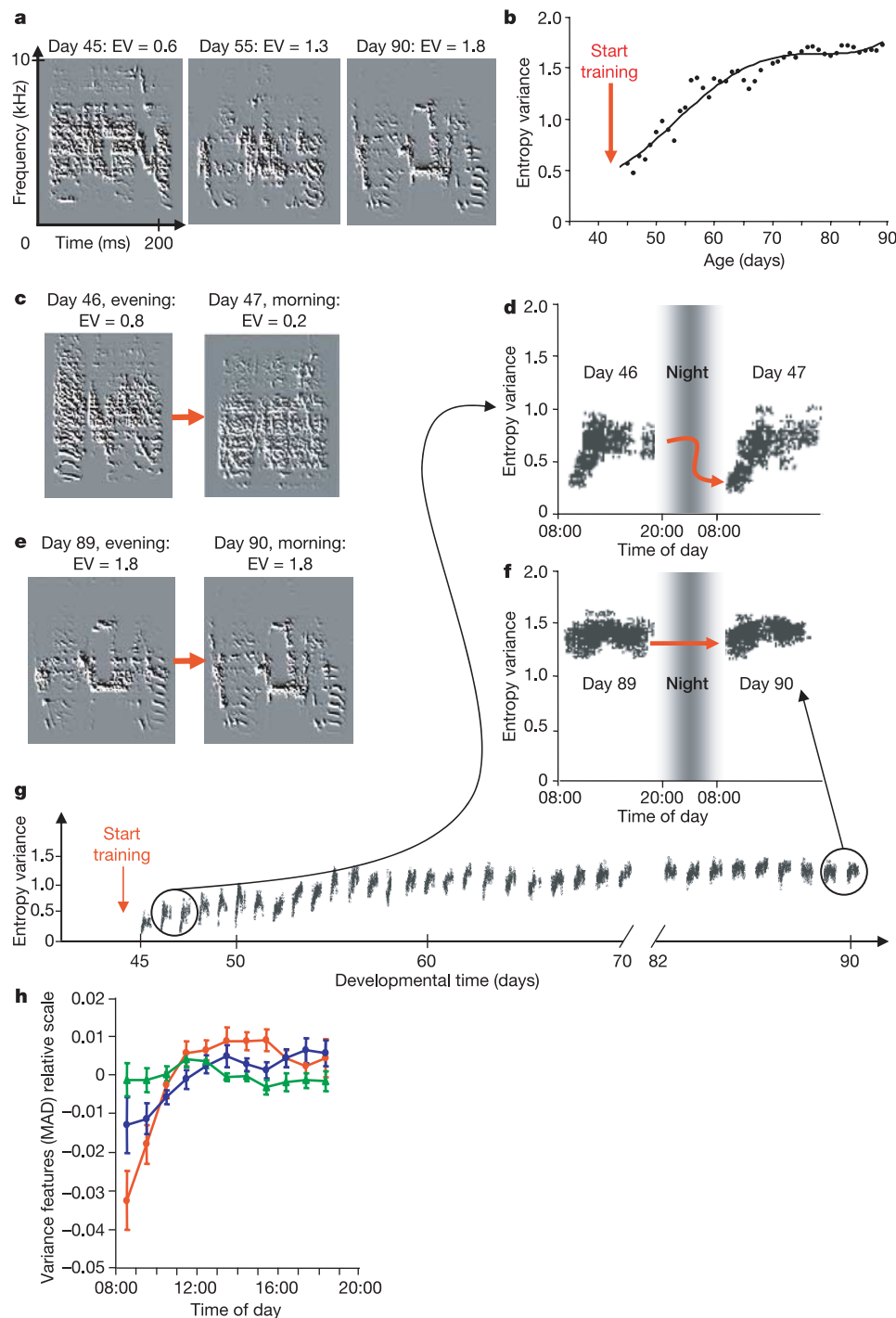


Figure 3 Recovery of syllable structure during the morning. **a**, Three examples of increasing syllable structure during development. **b**, Developmental change in the structure of the same syllable as captured by Wiener EV. **c**, Changes in the structure of that syllable during night-sleep. **d–f**, Tracking EV values continuously shows a decrease in EV values after the night-sleep of day 46 (**d**) but not after the night-sleep of day 89 (**e, f**). **g**, Tracing EV values continuously during development shows oscillations between days 45 and 60. EV values have been smoothed with a running median (period = 40 data

points). **h**, Collapsing together all variance features across birds shows the time course of syllable structure recovery during the day. Red curve, early syllable development (days 50–60); blue curve, late syllable development (days 85–90); green curve, 1-year-old birds. Features units were transformed to median absolute deviation from the mean (MAD). Mean daily MAD values are shifted to zero, to enhance the display of hourly changes in syllable structure. Error bars are s.e.m.

not correlate with better imitation (Spearman $\rho = 0.25$, $P = 0.443$), further confirming that the initial acoustic distance from the song model does not predict song learning¹⁷. In sum, the magnitude of daily oscillations in song structure, but not the total developmental change, seems to be a predictor of accurate song learning. A similar trend was observed in birds trained from day 90 (Supplementary Fig. 1).

We propose that the morning period of less structured (and perhaps more plastic) sounds gives the bird an opportunity to explore its vocal abilities and improve imitation. To examine the dynamic relations between daily changes in song structure and the progression of vocal learning, we selected the most complex syllable of each bird's song and computed similarity to the model syllable continuously (for each syllable produced) from day 55 to day 70 after hatching. As expected, similarity values dropped during night-sleep (repeated-measures ANOVA, $P < 0.001$) and then increased to a plateau (Fig. 4b). To measure a bird's overall progress towards imitating the tutor song, we defined a measure called 'record similarity', the quality of a bird's best success so far in copying the tutor song (Fig. 4b, c). For robustness, we computed the 95th centile of the similarities within a 1-h moving window. The record similarity at a given time is defined as the best 95th centile achieved so far. Improvements in record similarity are computed with reference to developmental time, not to the previous hour, thus excluding improvements that are due solely to recovery (Fig. 4b, c).

Strong improvements in record similarity were achieved during

the third hour of the photophase (Fig. 4d), overlapping with the late phase (third hour) of the recovery of song structure (Fig. 3h, red curve). Little improvement was achieved during the second half of the day. This pattern was not a consequence of different rates of singing across the day, because dividing the increases in record similarity by the singing rate (Fig. 4e) gave similar results, with large increases in record similarity from 3 to 6 h and very low increases during the second half of the day. Once syllable structure reached an asymptote at midday, record similarity did not increase much further, despite continued singing.

Causes of post-sleep deterioration

The underlying cause of the observed effects might be sleep, but there are alternative hypotheses. First, sleep inertia, a transitional state of lowered arousal³⁴, and possibly altered states of the vocal apparatus ('morning voice'), could explain the deterioration in song structure. Alternatively, vocal changes might be caused by circadian changes in hormonal state. To test both hypotheses, we prevented the bird from singing for 2 h during one morning ($n = 6$, range: 50–57 days) by taking the cage from the sound box into the laboratory³⁵. The bird did not attempt to sing but exhibited usual activities including eating and calling. When singing resumed (3–4 h after lights were turned on) we would have expected to see at

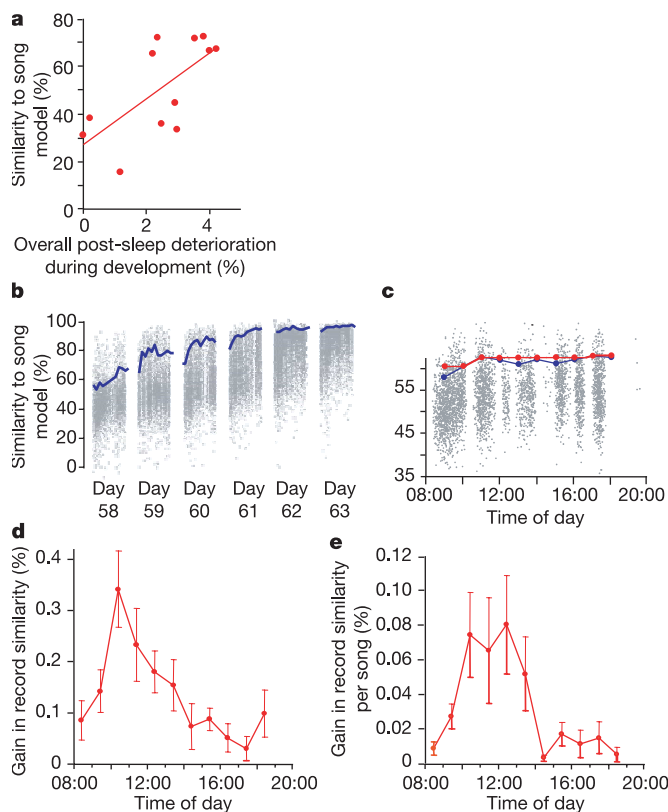


Figure 4 Progression of song learning. **a**, Relation between magnitude of post-sleep deterioration and similarity to the song model. **b**, An illustration of similarity measurements during six consecutive days. Grey dots represent the raw similarity scores and the blue curve shows the 95th centile. **c**, An illustration of record similarity measurements during one day. Red segments illustrate record similarity values and the blue curve shows the 95th centile. **d**, Gain in record similarity to the song model during the day across birds (means $\pm$ s.e.m.). Similarity scores were computed continuously from day 55 to day 70 in 12 birds. **e**, As in **d**, but divided by the number of songs produced in each 1-h time bin (means $\pm$ s.e.m.).

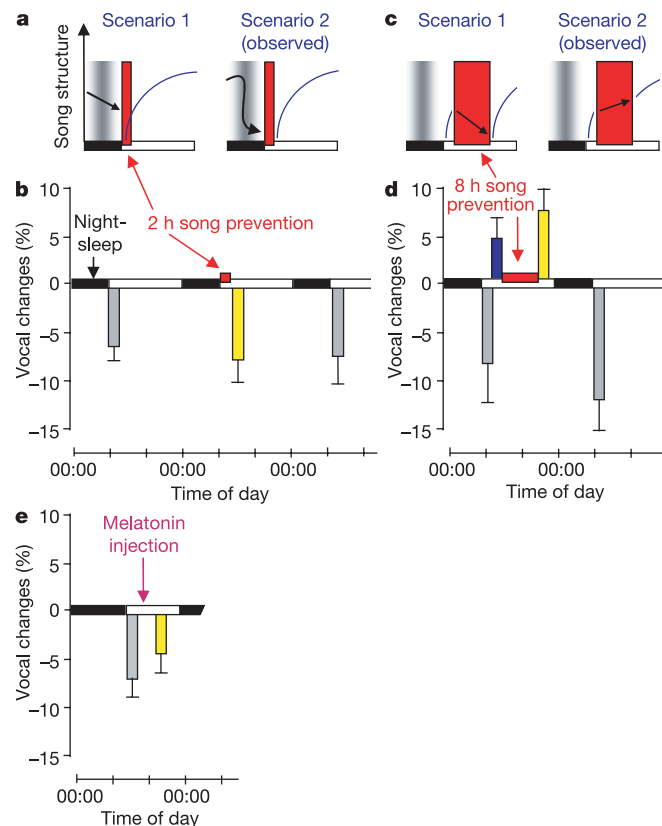


Figure 5 Vocal changes after song prevention and sleep manipulation. **a**, Recovery of song structure could be circadian (scenario 1) or it might require singing activity (scenario 2). **b**, Vocal changes (in reference to the developmental trend) were measured after night-sleep plus 2 h of song prevention (yellow bar, mean $\pm$ s.e.m.). Results were similar to those of post-sleep deterioration alone (grey bars, n.s., $n = 6$ birds). **c**, If deterioration of song structure during sleep were due to lack of practice, 8 h of song prevention should cause deterioration (scenario 1) or the bird might continue the morning recovery when singing resumed (scenario 2). **d**, Vocal changes during 8 h of song prevention (yellow bar) were similar to morning changes (from 08:00 to 10:00, blue bar, n.s.) and differed from vocal changes during night-sleep (grey bars, $P < 0.02$, $n = 9$ birds). **e**, Vocal changes after melatonin-induced sleep ($n = 6$ birds). Results in **b–e** are means and s.e.m.

least some recovery of song structure if the cause had been sleep inertia or circadian hormonal state (Fig. 5a), but song deterioration remained similar to that observed the day before (Fig. 5b, Wilcoxon, n.s.). Thus, post-sleep deterioration is not simply a circadian phenomenon and cannot be explained by sleep inertia.

Second, it could be that maintaining song structure during days of rapid learning requires intense practice, and that the lack of singing practice during sleep causes deterioration in motor performance. To test this hypothesis, we raised birds under a light regime of 16 h light and 8 h of darkness (16h:8h LD) to allow matching of the 8-h night-sleep with an 8-h song prevention interval during the photophase ($n = 9$, range 50–57 days). We monitored song development on-line to detect the time of rapid learning, and confirmed that post-sleep deterioration took place after 8 h of darkness. In the morning we allowed the bird to partly recover its song structure for 2 h. Then we prevented singing for 8 h; finally we allowed the bird to restart singing during the evening of the same day. When singing resumed, we would have expected to see deterioration if lack of practice had an effect, and continued improvement otherwise (as would be expected when comparing the second to the third hour of singing in the morning; Fig. 5c). As shown, vocal changes continued in a positive direction after the 8 h of song prevention (Fig. 5d). The difference between the effects of night-sleep and those of song prevention was statistically significant (Wilcoxon, $P = 0.02$).

Finally, we tested for causal relations between sleep and deterioration of song structure by inducing sleep during the day. Melatonin can induce sleep in zebra finches, and electrophysiological recordings suggest that song replay occurs readily in this induced sleep state²⁷. We allowed each bird ($n = 6$, age 54–63 days) to recover its song structure for 4 h after night-sleep and then injected 3 μ g of melatonin²⁷. All birds fell asleep within 15 min, and slept for 2–3 h during 4 ± 0.5 h of vocal rest. When singing resumed we observed song deterioration in all birds, comparable to that observed after night-sleep (Fig. 5e, Wilcoxon, n.s.).

Vocal changes during sleep could relate to synaptic²¹ or cellular^{36,37} remodelling that might occur during sleep. Regardless of the specific mechanism, our results indicate the involvement of an active process, perhaps neural song-replay during sleep. How might the bird's 'internal singing'^{26,27} during sleep give rise to the observed plasticity? If the lack of auditory feedback during replay during sleep has a similar effect to that of abolishing^{32,33,38} or perturbing³⁹ auditory feedback, we would expect to see drifts in song structure after sleep replay. We analysed post-deafening deterioration in eight adult birds and found that the deterioration of intra-syllabic temporal structure was similar to that observed after night-sleep in young birds (Fig. 6). Therefore, song replay along with lack of auditory feedback could by itself explain the decrease in song

structure after prolonged sleep. If feedback-guided singing never resumes (that is, after deafening), this effect is detrimental, but in the intact bird, transitory drifts in song structure could promote learning.

Discussion

Many procedural and declarative learning skills are enhanced after sleep^{1–11} with no apparent oscillations in performance. Oscillations in performance have been reported in the behaviour of several species, but possible relations to sleep were not investigated. For example, bumblebees can learn to approach and probe unfamiliar flowers faster as they gain experience during a foraging day, but these skills deteriorate strongly overnight⁴⁰. Another example is the development of voluntary movement in vertebrates. Movement patterns appear gradually in the newborn infant, progressing from simple to complex, but after prolonged immobility, movements become simple and primitive⁴¹ ('warm up' phenomenon). A similar effect was observed in the development of exploratory behaviour in the juvenile rat⁴².

Our study bridges these past observations by demonstrating a sleep-associated oscillation in performance that has a role in developmental learning. A potential reason why such a phenomenon has not been reported before is that the effect is specific to (or stronger in) developmental learning. A second reason for this might be the requirement for exhaustive, long-term behavioural measurements spanning many days to establish the effect.

We advance some speculations about the utility of these sleep-related oscillations in performance. Our speculations are based on the observation that a larger oscillation is associated with better final imitation. The oscillations could reflect a compromise between plasticity and consolidation of structure: periodic increases in plasticity might allow improvement of similarity to the song model through the correction of inappropriately consolidated structure.

Finally, it is worth noting that certain optimization algorithms such as simulated tempering⁴³ use non-monotonic trajectories of parameters. The ideas of simulated annealing⁴⁴ and tempering⁴³ come from metallurgical processes such as the reduction of brittleness of steel through tempering, which involves a cyclical process of reheating and cooling. □

Methods

Animal care

All experiments were performed in accordance with guidelines of the National Institutes of Health and have been reviewed and approved by the Institutional Animal Care and Use Committee of the City College of New York and the Wesleyan University.

Experimental design

We used 50 zebra finches (*Taenyopygia guttata*) from the City College of New York breeding colony. Colony management and experimental design have been described previously¹⁷. All birds were kept in isolation from days 30 to 90 after hatching.

Recordings from deafened birds were obtained from eight birds from the breeding colony of the Wesleyan University. These birds were kept in a large aviary with their parents and siblings of both sexes until days 70–90, when they were removed and housed together in groups of two to five.

Experimental groups

Training from day 43 to day 90: 12 birds from 8 families were trained with song playbacks¹⁷, starting from day 43 after hatching. We used three different songs (four birds per song model). Birds were raised from hatching under an artificial photoperiod of 12 h:12 h LD.

Training from day 60 to day 90: six birds from three different families were trained from day 60 with the three song models used in the first group (12 h:12 h LD).

Training from day 90 to day 120: eight birds from seven different families were trained from day 90 with the three song models used in the first group (12 h:12 h LD). Birds stayed in acoustic isolation until day 120.

No training: six birds from four different families were kept in similar sound boxes from day 30 to day 90 but were not trained (12 h:12 h LD).

Singing prevention: in 15 birds from 10 families, photoperiod was shifted from 12 h:12 h LD (days 0–30) to 16 h:8 h LD thereafter. Birds were trained from day 43 to day 90 (five birds per song model). We prevented singing once in each bird by taking the bird (with its cage) out of the training box and placing it in the laboratory under

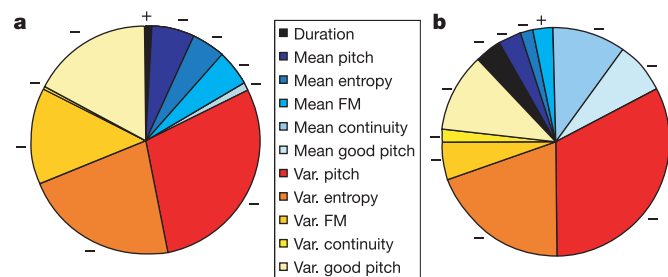


Figure 6 Comparison of post-sleep deterioration and post-deafening deterioration. **a**, Post-sleep changes in birds trained from day 43. **b**, Post-deafening changes (2–4 weeks after deafening). Each slice indicates the mean relative contribution of each feature to the overall effect. The sign indicates whether the feature value increased or decreased. Var., variance.

observation³⁵. We found that the young bird did not attempt to sing when moved to a new location but performed activities such as perch jumping, eating and frequent long and short calls, suggesting that the stress level was not excessive. We detected vocal changes by tracking song development on-line with the DVD-maps module of Sound Analysis Pro³¹ (<http://ofer.sci.ccnycunyc.edu>). Singing was prevented on a single day when rapid vocal changes were observed (between day 50 and day 57). In six birds we prevented singing during the first 2 h of the photophase, and in nine birds we prevented singing for 8 h, from the third to the tenth hour of the photophase (the same duration as the night in this group). In practice, however, singing resumed only about 2 h after our manipulation (2 h song prevention, 99 ± 24 min (mean $\pm$ s.e.m.), $n = 6$; 8 h song prevention, 121.6 ± 26.4 min, $n = 9$).

Tutoring prevention: to test the effect of training regimen, we halted training with song playbacks for 2 h during a single day of training (chosen randomly between day 50 and day 55) in five birds trained from day 43 to day 90.

Live tutoring: a young bird was raised and isolated as described previously (12 h:12 h LD). On day 43, an adult male (more than 1 year old, without any affiliation with the pupil) was introduced into its cage and both males stayed together until day 90. To obtain sound data from the young bird, we had to develop a custom source separation algorithm. We had to resolve the problem of source separation because the bird performed a close imitation of its tutor.

Melatonin-induced sleep: six birds were trained from day 43 to day 90 (one song model; 12 h:12 h LD). We confirmed deterioration after night-sleep during that morning (age varied between 54 and 63 days after hatching), and on the fourth hour of the photophase we injected 3 μ g of melatonin (Calbiochem) in phosphate-buffered saline (0.3 ml) intraperitoneally. Behavioural state was measured by continuous video recording (see 'Observations on sleep' below).

Deafened birds: eight birds were deafened by bilateral cochlear removal¹⁹ between 123 and 140 days after hatching. Songs were recorded 2–4 weeks after deafening.

Observations on sleep

The zebra finch is a diurnal species that does not sing in darkness. Infrared video recording confirmed that juvenile birds (aged 50–60 days, $n = 5$) presented characteristic postures of sleep^{26,28} (that is, head under wing and deep slow breathing) during most of the night. The same behavioural criteria were used to assess melatonin-induced sleep.

Song recording, storage and initial analysis

Using Sound Analysis Pro³¹ we recorded the entire song development in all birds, with the exception of occasional system crashes that caused the loss of up to 10% of the raw data. Song activity was detected automatically and saved (16 bits, sampling frequency 44.1 kHz) continuously throughout the experiment. We recorded and analysed 2.2 terabytes of song, stored as wave files in a DeskForce II tera-storage raid system and on DVDs. Songs were analysed with the batch module of Sound Analysis Pro, and results (for example, syllable features) were stored in MySQL 4.0 tables (<http://mySQL.com>). Final stages of analysis were performed with Microsoft Excel and MATLAB (The Mathworks, Natick, MA).

Spectral analysis and computation of acoustic features

We performed multitaper spectral analysis⁴⁵ to compute spectral derivatives and acoustic features as documented in the Sound Analysis Pro user manual. Subsequent analysis was based on the six acoustic features computed on each spectral frame: amplitude, pitch, entropy, FM, continuity and goodness of pitch³⁹.

Methods of segmentation into syllables, computing syllable features, clustering syllables and tracing the evolution of clusters, estimating vocal changes across time scales, similarity measurements and hypothesis testing are presented in Supplementary Methods with appropriate references.

Received 29 October; accepted 15 December 2004; doi:10.1038/nature03275.

- Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. & Sagi, D. Dependence on REM sleep of overnight improvement of a perceptual skill. *Science* **265**, 679–682 (1994).
- Wilson, M. A. & Mc Naughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994).
- Maquet, P. The role of sleep in learning and memory. *Science* **294**, 1048–1052 (2001).
- Peigneux, P., Laureys, S., Delbecq, X. & Maquet, P. Sleeping brain, learning brain. The role of sleep for memory systems. *Neuroreport* **12**, 111–124 (2001).
- Stickgold, R., Hobson, J. A., Fosse, R. & Fosse, M. Sleep, learning, and dreams: off-line memory reprocessing. *Science* **294**, 1052–1057 (2001).
- Fenn, K. M., Nusbaum, H. C. & Margoliash, D. Consolidation during sleep of perceptual learning of spoken language. *Nature* **425**, 614–616 (2003).
- Walker, M. P., Brakefield, T., Hobson, A. & Stickgold, R. Dissociable stages of human memory consolidation and reconsolidation. *Nature* **425**, 616–620 (2003).
- Wagner, U., Gals, S., Halder, H., Verleger, R. & Born, J. Sleep inspires insight. *Nature* **427**, 352–355 (2004).
- Walker, M. P. & Stickgold, R. Sleep-dependent learning and memory consolidation. *Neuron* **44**, 121–133 (2004).
- Huber, R., Ghilardi, M. F., Massimini, M. & Tononi, G. Local sleep and learning. *Nature* **430**, 78–81 (2004).
- Peigneux, P. *et al.* Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron* **44**, 535–545 (2004).
- Frank, M. G., Issa, N. P. & Stryker, M. P. Sleep enhances plasticity in the developing visual cortex. *Neuron* **30**, 275–287 (2001).
- Zuckerman, B., Stevenson, J. & Bailey, V. Sleep problems in early childhood: continuities, predictive

- factors, and behavioral correlates. *Pediatrics* **80**, 664–671 (1987).
- Doupe, A. S. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
- Brainard, M. S. & Doupe, A. J. What songbirds teach us about learning. *Nature* **417**, 351–358 (2002).
- Immelmann, K. in *Bird Vocalizations* (ed. Hinde, R. A.) 61–74 (Cambridge Univ. Press, Cambridge, 1969).
- Tchernichovski, O., Mitra, P. P., Lints, T. & Nottebohm, F. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* **291**, 2564–2569 (2001).
- Tchernichovski, O. & Mitra, P. P. Towards quantification of vocal imitation in the zebra finch. *J. Comp. Physiol. [A]* **188**, 867–878 (2002).
- Konishi, M. The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Z. Tierpsychol.* **22**, 770–783 (1965).
- Doupe, A. J. Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J. Neurosci.* **17**, 1147–1167 (1997).
- Botter, S. W. & Johnson, F. Circuits, hormones, and learning: vocal behavior in songbirds. *J. Neurobiol.* **33**, 602–618 (1997).
- Schmidt, M. F. & Konishi, M. Gating of auditory responses in the vocal control system of awake songbirds. *Nature Neurosci.* **1**, 513–518 (2001).
- Nick, T. A. & Konishi, M. Dynamic control of auditory activity during sleep: correlation between song response and EEG. *Proc. Natl Acad. Sci. USA* **98**, 14012–14016 (2001).
- Rauske, P. L., Shea, S. D. & Margoliash, D. State and neuronal class-dependent reconfiguration in the avian song system. *J. Neurophysiol.* **89**, 1688–1701 (2003).
- Nick, T. A. & Konishi, M. Neural song preference during vocal learning in the zebra finch depends on age and state. *J. Neurobiol.* **62**, 231–242 (2005).
- Dave, A. S. & Margoliash, D. Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* **290**, 812–816 (2000).
- Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).
- Margoliash, D. Evaluating theories of bird song learning: implications for future directions. *J. Comp. Physiol. A* **188**, 851–866 (2002).
- Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B. & Mitra, P. P. A procedure for an automated measurement of song similarity. *Anim. Behav.* **59**, 1167–1176 (2000).
- Tchernichovski, O., Lints, T., Derégnaucourt, S., Cimenser, A. & Mitra, P. P. Studying the song development process: rationale and methods. *Ann. NY Acad. Sci.* **1016**, 348–363 (2004).
- Tchernichovski, O. & Mitra, P. P. Sound analysis Pro user manual. (<http://ofer.sci.ccnycunyc.edu>) (2004).
- Lombardino, A. J. & Nottebohm, F. Age at deafening affects the stability of learned song in adult male zebra finches. *J. Neurosci.* **20**, 5054–5064 (2000).
- Nordeen, K. W. & Nordeen, E. J. Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav. Neural Biol.* **57**, 58–66 (1992).
- Tassi, P. & Muzet, A. Sleep inertia. *Sleep Med. Rev.* **4**, 341–353 (2000).
- Jarvis, E. D. & Nottebohm, F. Motor-driven gene expression. *Proc. Natl Acad. Sci. USA* **94**, 4097–4102 (1997).
- Nordeen, K. W. & Nordeen, E. J. Projection neurons within a vocal motor pathway are born during song learning in zebra finches. *Nature* **334**, 149–151 (1988).
- Wang, N., Hurley, P., Pytte, C. & Kirn, J. R. Vocal control neuron incorporation decreases with age in the adult zebra finch. *J. Neurosci.* **22**, 10864–10870 (2002).
- Brainard, M. S. & Doupe, A. J. Postlearning consolidation of birdsong: stabilizing effects of age and anterior forebrain lesions. *J. Neurosci.* **21**, 2501–2517 (2001).
- Leonardo, A. & Konishi, M. Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature* **399**, 466–470 (1999).
- Kearar, M., Motro, U., Shur, Y. & Schmida, A. Overnight memory retention of foraging skills by bumblebees is imperfect. *Anim. Behav.* **52**, 95–104 (1996).
- Golani, I. A mobility gradient in the organization of vertebrate movement: the perception of movement through symbolic language. *Behav. Brain Sci.* **15**, 249–308 (1992).
- Tchernichovski, O., Benjamini, Y. & Golani, I. The dynamics of long-term exploration in the rat. Part I. A phase-plane analysis of the relationship between location and velocity. *Biol. Cybern.* **78**, 423–432 (1998).
- Marinari, E. & Parisi, G. Simulated tempering: a new Monte Carlo scheme. *Europhys. Lett.* **19**, 451–458 (1992).
- Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
- Ho, C. E., Pesaran, B., Fee, M. S. & Mitra, P. P. in *Proc. Joint Symposium on Neural Computation* Vol. 5, 76–83 (Univ. California Press, San Diego, 1998).
- Loader, C. *Local Regression and Likelihood* (Springer, New York, 1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Swigger for programming; J. Wallman for comments; P. Andrews for help with graphics; and T. Lints, K. Maul, A. Alexander and L. Dayani for help in analysing data. This work was supported by grants from the NIH to O.T. and P.P.M., by a NIH RCMI (National Institutes of Health Research Centers in Minority Institutions) grant to the City College of New York, by internal support from the Cold Spring Harbor Laboratory and the Robertson foundation to P.P.M., and from the Fondation Fyssen to S.D.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.D. (sderegna@sci.ccnycunyc.edu).

Morphological differences between Saturn's ultraviolet aurorae and those of Earth and Jupiter

J. T. Clarke¹, J.-C. Gérard², D. Grodent², S. Wannawichian¹, J. Gustin², J. Connerney³, F. Cray⁴, M. Dougherty⁵, W. Kurth⁶, S. W. H. Cowley⁷, E. J. Bunce⁷, T. Hill⁸ & J. Kim⁹

¹Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

²Université de Liège, Allée du 6 Aout - Sart Tilman, B4000 Liège, Belgium

³NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

⁴Southwest Research Institute, Culebra Road, San Antonio, Texas 78228, USA

⁵Imperial College, London SW7 2AZ, UK

⁶University of Iowa, Iowa City, Iowa 52242, USA

⁷University of Leicester, Leicester LE1 7RH, UK

⁸Rice University, Houston, Texas 77005, USA

⁹School of Civil and Environmental Engineering, Yonsei University, Seoul 120-749, Korea

It has often been stated that Saturn's magnetosphere and aurorae are intermediate between those of Earth, where the dominant processes are solar wind driven¹, and those of Jupiter, where processes are driven by a large source of internal plasma^{2–4}. But this view is based on information about Saturn that is far inferior to what is now available. Here we report ultraviolet images of Saturn, which, when combined with simultaneous Cassini measurements of the solar wind⁵ and Saturn kilometric radio emission⁶, demonstrate that its aurorae differ morphologically from those of both Earth and Jupiter. Saturn's auroral emissions vary slowly; some features appear in partial corotation whereas others are fixed to the solar wind direction; the auroral oval shifts quickly in latitude; and the aurora is often not centred on the magnetic pole nor closed on itself. In response to a large increase in solar wind dynamic pressure⁵ Saturn's aurora brightened dramatically, the brightest auroral emissions moved to higher latitudes, and the dawn side polar regions were filled with intense emissions. The brightening is reminiscent of terrestrial aurorae, but the other two variations are not. Rather than being intermediate between the Earth and Jupiter, Saturn's auroral emissions behave fundamentally differently from those at the other planets.

Geomagnetic activity and brightenings of the Earth's auroral emissions are controlled primarily by the interaction of the solar wind with the Earth's magnetic field. Geomagnetic substorms occur during intervals of southward interplanetary magnetic field (IMF), permitting reconnection between the IMF and geomagnetic fields. Substorms are manifested as bright auroral emissions, which typically shift the brightened auroral oval to lower latitudes and also fill in the midnight sector of the oval¹. The Earth's auroral oval is adjacent to and equatorward of the boundary between open and closed magnetic field lines, and its statistical location is fixed with respect to the solar wind direction rather than rotating with the planet. By contrast², Jupiter's main auroral oval occupies field lines that extend only to the middle magnetosphere, where outward diffusing plasma lags behind the rotational speed of Jupiter, driving strong currents in and out of Jupiter's ionosphere^{3,4}. Jupiter's total auroral power is relatively constant, rarely changing by as much as a factor of two, whereas the localized auroral storms that are observed consist of rapid flares poleward of the main oval and more gradual brightenings of the dawn sector of the main oval.

The brightness of Saturn's auroral emissions, the size of its magnetosphere, and its magnetic moment all exceed those at the Earth and are less than those at Jupiter (Table 1). Because

Saturn's auroral oval appears at higher magnetic latitudes than Jupiter's, it has long been believed⁷ to map close to the boundary of open and closed field lines, and on theoretical grounds⁸ it has been considered more likely to be controlled by the solar wind than Jupiter's. At the same time, Saturn's magnetosphere is much larger than the Earth's (Table 1), and Saturn's rapid rotation and internal plasma sources suggest that jovian-like processes should also be active. The approach of the Cassini spacecraft to Saturn has provided a unique opportunity to examine the nature of Saturn's auroral emissions. A series of five Hubble Space Telescope (HST) orbits on 8 January 2004 provided measurements of the longitude variations of the auroral emissions over 70% of a Saturn rotation, while single orbit observations on 12 other days were obtained to compare the auroral activity with varying solar wind conditions. Twelve image sub-frames are shown in Fig. 1, obtained with the HST Space Telescope Imaging Spectrograph (STIS).

From the complete set of observations, it has become clear that Saturn's auroral emissions vary on timescales down to tens of minutes, similar to Jupiter's main oval but far slower than its polar emissions. The images from 8 January (Fig. 1) show the presence of a localized bright emission region along the main oval at 130–170° λ_{SLS} (Saturn longitude system), rotating in the direction of the planet, becoming fainter with time, and shifting to higher latitudes. This emission region may correspond to earlier reports of an active sector in both the ultraviolet and Saturn kilometric radiation (SKR) emissions based on Voyager measurements⁷, and might correspond to the source of SKR emission⁶. The brightest ultraviolet emissions (concentrating on emissions that appear stable over time) move at 30–65% of Saturn's rotation speed, whereas the motion of the overall emission pattern appears to be best fitted by 75% of the corotation rate. This would imply that the emissions map along Saturn's magnetic field to a region where the plasma is moving at 30–75% of the corotation rate in Saturn's magnetic field. As the plasma generally rotates at a lower angular velocity with increasing distance, the rotational motions of auroral emissions can be compared with plasma flow speeds to estimate the equatorial distance from the planet connecting to these auroral emissions. From earlier Voyager measurements of the plasma flow⁹, this could be consistent with a wide range of distances from the planet, from 6–7 Saturn radii (R_S) out to the magnetopause boundary. Plasma flow measurements from Cassini orbiting Saturn, in comparison with auroral images, will be helpful in determining the distance to which the main auroral oval maps. This observation of partially corotating emissions contrasts with earlier HST images of Saturn's auroral emissions^{10–12}, which indicated that the emissions consistently brightest in the dawn sector—that is, fixed in local time. This raises the possibility of different states of Saturn's magnetosphere, leading to different auroral emission distributions. Such states could be related to the orientation of Saturn with respect to, or variations in, the solar wind and IMF.

The general distribution of the auroral emissions is not what was expected from the known axisymmetric magnetic field and internal plasma environment of Saturn. The auroral oval was thought to be symmetric about the rotational and magnetic axis. It was also expected to map to near the open/closed field line boundary and be longitudinally continuous, with possible local time variations in intensity reflecting the solar wind's interaction with Saturn's magnetosphere. The present observations show the entire southern oval, owing to the 26° tilt of Saturn at this epoch. To determine the central location and size of the auroral oval, a least squares fit has been made of a circular auroral oval, with variable radius and centre, to the brightest emissions in the oval for each day. This reveals an offset averaging 3–4° toward midnight in 10 of the 17 observing periods, similar to what is seen in the Earth's auroral emissions. This may be understood as simply reflecting the

Table 1 Parameters of magnetospheres and auroral emissions

Planet	Magnetic moment/ field strength	Magnetosphere cross-section diameter	Solar wind travel time, bow to planet	Average auroral brightness/ input power
Earth	1/0.3 G	3×10^5 km	3 min	1–100 kR/1–100 $\times 10^9$ W
Jupiter	20,000/4 G	200×10^5 km	200 min	10–1,000 kR/a few $\times 10^{13}$ W
Saturn	580/0.2 G	70×10^5 km	50 min	1–100 kR/1–10 $\times 10^{11}$ W

Field strengths are average values at the equator, and magnetic moments are scaled to 1 for the Earth. Travel times are listed for an assumed solar wind speed of 400 km s^{-1} . Average auroral emission brightnesses are measured, whereas the input power is determined by detailed modelling of energy deposition in the atmosphere as related to the product of brightness and emitting area. Initial observations of Saturn's magnetosphere and aurora were carried out by the Pioneer 11¹⁵ and Voyager 1 and 2 spacecraft¹⁴ over 1979–81. In the 1980s there were occasional measurements of strong auroral H Ly α emission brightenings by the IUE satellite^{15,16}, and in the past decade about 30 high resolution images of Saturn's ultraviolet auroral emissions were obtained with the HST, both the Wide Field and Planetary Camera 2¹⁰ and STIS^{11,12}. There has been one other attempt to relate Saturn's auroral intensity to the solar wind based on two images and a 9 AU extrapolation of the solar wind from the Earth to Saturn¹⁷.

solar wind's effect of compressing the dayside magnetosphere and expanding the nightside magnetotail, shifting the dayside auroral oval to higher latitudes and the nightside oval to lower latitudes. By contrast, the partial corotation of the emission features is similar to Jupiter's main aurora oval, which maps to the middle magnetosphere region of sub-corotating plasma. The broken appearance of the oval on several days and latitudinal motions of several degrees are both pronounced and surprising, and may initially be interpreted in terms of the response to variations in the solar wind pressure or perhaps to reconnection-related phenomena. These motions, much larger than seen on Jupiter, indicate a more dynamic environment than expected in Saturn's magnetosphere.

One intriguing feature in Saturn's auroral oval is an apparent spiral pattern with local time. The brightest emissions in the dawn sector, when followed through noon, dusk and midnight, frequently returned to the dawn sector much fainter and at lower latitudes, not connecting to the brighter start of the oval. This is reminiscent of Io's footprint emission and downstream trail in Jupiter's aurorae, but an analogous phenomenon at Saturn would require a source of plasma near the open/closed field line boundary. Saturn's satellite Titan orbits at the required distance, but the spiral pattern appeared fixed in local time while Titan moved through more than one complete orbit during the period of imaging. Furthermore, Saturn auroral images from January 2001 showed a spiral pattern having the same rotational sense but with the disconnected region near

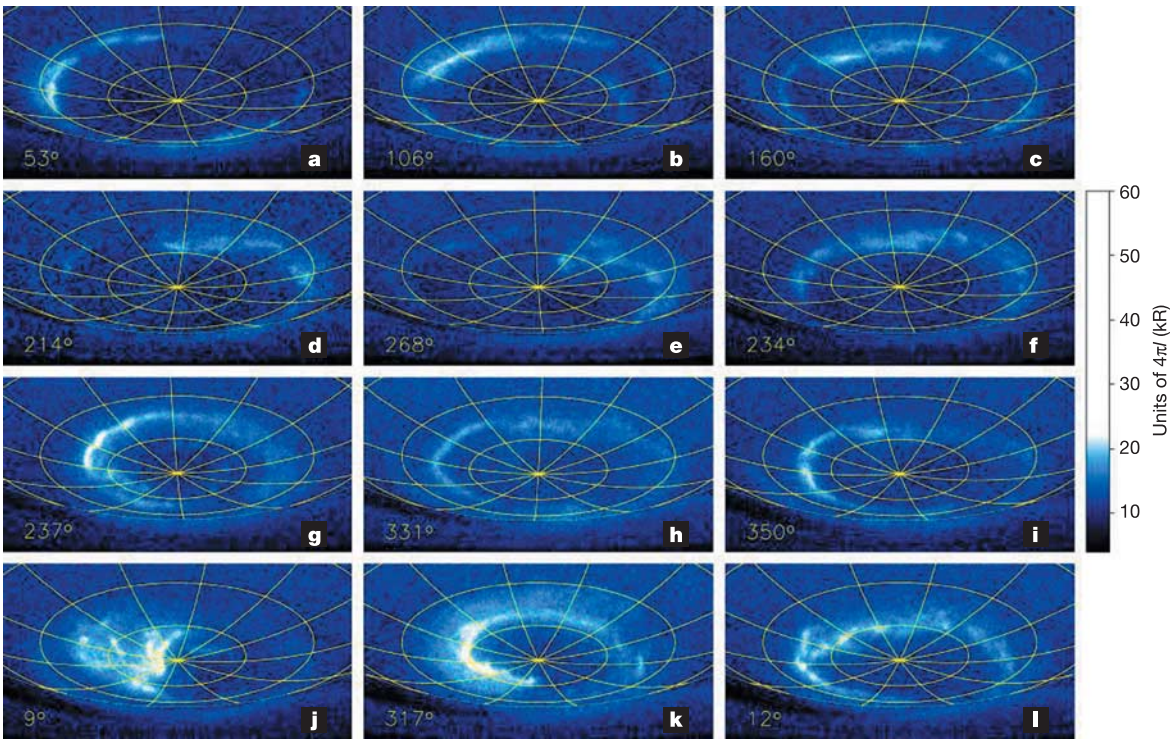


Figure 1 Ultraviolet images of Saturn's southern aurora from HST STIS. The first five frames are images from five HST orbits on 8 January 2004 (a–e), followed by images from 14 (f), 18 (g), 23 (h), 24 (i), 26 (j), 28 (k) and 30 January 2004 (l). Saturn north is up in each frame, with latitude and longitude lines plotted every 10° and 30°. Numbers in each panel are the central meridian longitude in degrees in the Saturn longitude system (SLS). The 26° tilt of Saturn's southern pole towards the Earth makes the entire auroral region visible. The planet and rings are seen in reflected ultraviolet sunlight, while the auroral emissions about the south pole are excited by precipitating energetic electrons. Each frame combines two clear images (total 540 s exposure) taken in one HST orbit,

with a logarithmic stretch of the surface brightness in kilorayleighs (kR). The limiting sensitivity and the brightest emissions are approximately 1 and 100 kR, respectively. The images have a bandpass of 115–174 nm, with sensitivity to the auroral spectrum including the H₂ Lyman (B¹Σ_u⁺–X¹Σ_g⁺) band emissions and part of the Werner (C¹Π_u–X¹Σ_g⁺) band series plus the H Lyman-α line. The pixel size of 0.024 arcsec provides reasonable sampling of the point spread function of 0.08 arcsec (or 490 km on Saturn), but the imaging resolution is further degraded by Saturn's rotation of roughly 1° per 100 s. *I*, intensity

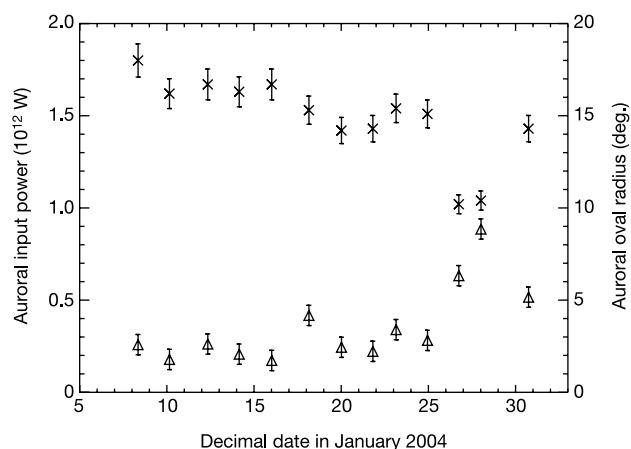


Figure 2 Variations in Saturn's auroral power and radius of the oval. Saturn's total southern auroral input power (triangles; left-hand y axis) for each day of observations in January 2004 is overplotted with the mean radius (crosses; right-hand y axis) of the auroral oval. The oval radius has been determined from a least-squares fit of the brightest emissions to a circle. The total input power is estimated by adding all the polar emissions after subtraction of a simulated Saturn background disk and ring image, multiplying by 1.1 to account for ultraviolet emissions outside the STIS clear bandpass, then converting from photons produced to charged particle input energy, assuming that 1 mW m^{-2} of input energy corresponds to 10 kR in ultraviolet emission. This value is based on detailed modelling of the input energy deposition, interaction with the upper atmosphere, and radiative transfer of the departing emissions¹¹, which give a best fit mean energy for the incoming electrons of 12 keV. Error bars represent the counting statistics in the total auroral counts and the subtraction of the reflected solar background, and are dominated by the latter. The measurements of the radius of the auroral oval are average values from each day for the measured locations of the brightest emissions over a range of longitudes, with error bars indicating relative uncertainties from day to day from the least squares fit (standard deviation from the mean). An anti-correlation is seen between the auroral brightness or power and the radius of the auroral oval.

dusk rather than near dawn¹². This phenomenon will require further study. The apparent presence of an active region in a specific longitude range on 8 January 2004, perhaps due to a magnetic anomaly in that longitude sector⁶, implies some degree of rotational control and also deserves further study. If this were shown to be a long-lived feature of Saturn's auroral emissions at a fixed magnetic longitude, that would support its identification as a magnetic anomaly.

More detailed information on the response of Saturn's auroral emissions to the solar wind and IMF has been provided from two disturbances in the solar wind passing Saturn during the period of observations. On 18 January 2004 the auroral oval appeared brighter than average, with a reduced radius. These images were obtained about 40 hours after a solar wind shock had arrived at the bow of Saturn's magnetosphere⁵. A larger solar wind shock reached Saturn on 26 January⁵ about 11 hours before the images on that day shown in Fig. 1. This disturbance led to the most dramatic auroral brightenings on 26–28 January, when the auroral power increased by a factor of three from the mean over 8–24 January, and the dawn side oval completely filled in with bright emissions on 26 January. The morphology implies active precipitation over the entire dawn side of the polar cap. Overall, as seen in Fig. 2, the auroral power appears to have been anti-correlated with the radius of the auroral oval in January 2004, opposite to the trend observed at Earth, where the oval's brightness is positively correlated with its size.

Although the event of 26 January on Saturn bears some resem-

blance to the onset of an auroral substorm at Earth, terrestrial substorms usually begin in the midnight sector and last for tens of minutes, whereas Saturn's brightening was in the dawn sector and only slightly decreased over 50 minutes of observations on that day. Moreover, terrestrial substorms thicken the nightside oval both poleward and equatorward from its pre-substorm position, whereas at Saturn the brightening is restricted to latitudes poleward of the pre-existing oval.

Within the limits of non-continuous observations, in the images on 28 and 30 January 2004 the auroral oval appeared to fade and expand in radius on a timescale of several days. The long timescale contrasts sharply with substorm processes at the Earth, possibly reflecting the much larger scale size of Saturn's magnetosphere.

Comparison with measurements of the approaching solar wind from Cassini⁵ indicates that the auroral brightness and power correlate best with the solar wind dynamic pressure, rather than with any specific orientation of the IMF. Although solar wind driving is reminiscent of substorms on the Earth, the brightening of the dawn sector, the continued contraction of the oval, and the primary correlation with solar wind dynamic pressure rather than IMF direction are all characteristics that appear to be unique to Saturn. The latitudinal motions of the main oval, large response to solar wind changes, and filling in of the dawn side polar region are all in contrast with Jupiter. In addition, the polar flares observed on Jupiter have not been observed at Saturn. We conclude that among the three magnetospheres in the Solar System sufficiently well studied to be compared, the properties of auroral emissions at Saturn are unique to that planet. □

Received 17 August; accepted 22 December 2004; doi:10.1038/nature03331.

- Milan, S. E. *et al.* Variations in polar cap area during two substorm cycles. *Ann. Geophys.* **21**, 1121–1140 (2003).
- Hill, T. W. Auroral structures at Jupiter and Earth. *Adv. Space Res.* **33**, 2021–2031 (2004).
- Hill, T. W. The Jovian auroral oval. *J. Geophys. Res.* **106**, 8101–8107 (2001).
- Cowley, S. W. H. & Bunce, E. J. Origin of the main auroral oval in Jupiter's coupled magnetosphere-ionosphere system. *Planet. Space Sci.* **49**, 1067–1088 (2001).
- Crary, F. *et al.* Solar wind dynamic pressure and electric field as the main factors controlling Saturn's aurorae. *Nature* doi:10.1038/nature03333 (this issue).
- Kurth, W. *et al.* An Earth-like correspondence between Saturn's auroral features and radio emission. *Nature* doi:10.1038/nature03334 (this issue).
- Sandel, B. R. & Broadfoot, A. L. Morphology of Saturn's aurora. *Nature* **292**, 679–682 (1981).
- Cowley, S. W. H. & Bunce, E. J. Corotation-driven magnetosphere-ionosphere coupling currents in Saturn's magnetosphere and their relation to the auroras. *Ann. Geophys.* **21**, 1691–1707 (2003).
- Richardson, J. D. Thermal ions at Saturn: Plasma parameters and implications. *J. Geophys. Res.* **91**, 1381–1389 (1986).
- Trauger, J. T. *et al.* Saturn's hydrogen aurora: WFPC 2 imaging from the Hubble Space Telescope. *J. Geophys. Res.* **103**, 20237–20244 (1998).
- Cowley, S. W. H., Bunce, E. J. & Prangé, R. Saturn's polar ionospheric flows and their relation to the main auroral oval. *Ann. Geophys.* **21**, 1–16 (2003).
- Gérard, J.-C. *et al.* Characteristics of Saturn's FUV aurora observed with the Space Telescope Imaging Spectrograph. *J. Geophys. Res.* **109**, A09207, doi:10.1029/2004JA010513 (2004).
- Judge, D. L., Wu, F. & Carlson, R. Ultraviolet observations of the saturnian system. *Science* **207**, 431–434 (1980).
- Broadfoot, A. L. *et al.* Extreme ultraviolet observations from Voyager 1 encounter with Saturn. *Science* **212**, 206–211 (1981).
- Clarke, J. T. *et al.* IUE detection of bursts of H Ly- α from Saturn. *Nature* **290**, 226–227 (1981).
- McGrath, M. & Clarke, J. T. HI Lyman alpha emission from Saturn (1980–1990). *J. Geophys. Res.* **A 97**, 13691–13703 (1992).
- Prangé, R. *et al.* A CME-driven interplanetary shock traced from the Sun to Saturn by planetary auroral emissions. *Nature* (in the press).

Acknowledgements This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute (STScI), which is operated by AURA, Inc., for NASA. The research was supported by grants from STScI and NASA to Boston University. J.C.G. and D.G. acknowledge support from the Belgian Fund for Scientific Research and the PRODEX programme of ESA. S.W.H.C. was supported by a PPARC senior fellowship, and E.J.B. by a PPARC post-doctoral fellowship. F.C., W.K. and T.H. acknowledge support from the NASA/JPL Cassini project.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.T.C. (jclarke@bu.edu).

Solar wind dynamic pressure and electric field as the main factors controlling Saturn's aurorae

F. J. Crary¹, J. T. Clarke², M. K. Dougherty³, P. G. Hanlon³, K. C. Hansen⁴, J. T. Steinberg⁵, B. L. Barraclough⁵, A. J. Coates⁶, J.-C. Gérard⁷, D. Grodent⁷, W. S. Kurth⁸, D. G. Mitchell⁹, A. M. Rymer⁶ & D. T. Young¹

¹Southwest Research Institute, Culebra Road, San Antonio, Texas 78288, USA

²Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

³Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

⁴The University of Michigan, Space Research Building, Ann Arbor, Michigan 48109, USA

⁵Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

⁶University College London, Mullard Space Science Laboratory, Surrey RH5 6NT, UK

⁷Université de Liège, 17 avenue du 6 aout, B4000 Liège, Belgium

⁸Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242, USA

⁹Applied Physics Laboratory, Johns Hopkins University, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA

The interaction of the solar wind with Earth's magnetosphere gives rise to the bright polar aurorae and to geomagnetic storms¹, but the relation between the solar wind and the dynamics of the outer planets' magnetospheres is poorly understood. Jupiter's magnetospheric dynamics and aurorae are dominated by processes internal to the jovian system², whereas Saturn's magnetosphere has generally been considered to have both internal and solar-wind-driven processes. This hypothesis, however, is tentative because of limited simultaneous solar wind and magnetospheric measurements. Here we report solar wind measurements, immediately upstream of Saturn, over a one-month period. When combined with simultaneous ultraviolet imaging³ we find that, unlike Jupiter, Saturn's aurorae respond strongly to solar wind conditions. But in contrast to Earth, the main controlling factor appears to be solar wind dynamic pressure and electric field, with the orientation of the interplanetary magnetic field playing a much more limited role. Saturn's magnetosphere is, therefore, strongly driven by the solar wind, but the solar wind conditions that drive it differ from those that drive the Earth's magnetosphere.

In the case of Earth's auroral activity and associated magnetic substorms, the north-south component of the interplanetary magnetic field is especially important. A southward interplanetary magnetic field allows efficient dayside reconnection between the Earth's and the interplanetary magnetic field, and is an important driver of substorms, whereas a northward field results in high-latitude reconnection and significantly weaker control of magnetospheric dynamics by the solar wind¹. Solar wind shocks drive auroral activity, regardless of the interplanetary magnetic field orientation, and the magnitude of this effect correlates well with the solar wind speed and magnetic field strength. However, the influence of shocks is much more pronounced when the interplanetary magnetic field is southward⁴.

In contrast, Jupiter's rapid rotation and strong magnetic field, together with a large, internal plasma source from its satellite, Io, drive radial, outward transport of plasma and magnetospheric currents. These processes are the primary source of dynamics (ref. 2, and references therein) and the aurorae⁵. Despite this, the solar wind does exert some influence on Jupiter's magnetosphere⁶⁻⁹. On the basis of the size of the planet's magnetosphere and the observed ease with which its magnetopause is pushed inward and

outward by changing solar wind conditions, it has been suggested that dynamic pressure is the main solar wind driver, rather than the orientation of the interplanetary magnetic field^{10,11}. Saturn, like Jupiter, is a rapidly rotating planet and has significant internal plasma sources. However, its magnetic field and internal plasma sources are weaker than Jupiter's. Observations¹² by the Voyager spacecraft showed a significant degree of solar wind control, although the origin of these correlations was unclear. In the case of the Voyager studies, kilometric radio emissions were used as a proxy for magnetospheric dynamics.

Between 10 and 30 January 2004, the Cassini spacecraft measured the solar wind upstream of Saturn's magnetosphere while the Hubble Space Telescope (HST) made observations of Saturn's ultraviolet aurorae roughly once every other day³. The Cassini data discussed here come primarily from the CAPS instrument (Cassini Plasma Spectrometer Subsystem, providing ion and electron data from 1 eV to 50 keV; ref. 13) and from the magnetometer. During this period, Cassini solar wind ion measurements were largely continuous (81% duty cycle) and interrupted only by occasional downlinks of data to Earth. Cassini also made measurements of energetic particle flux, local plasma waves and auroral radio emissions from Saturn¹⁴.

During this period, the Cassini spacecraft was at a range of $(78 \pm 5) \times 10^6$ km from Saturn, over the post-dawn side of the planet, and $(31 \pm 2) \times 10^6$ km upstream. Although this distance seems large for an upstream solar wind monitor, the Saturn-Sun-spacecraft angle was under 3.3° . Studies of Cassini and Galileo data, obtained near Jupiter, indicate that a spacecraft may act as a viable upstream monitor when the planet-Sun-spacecraft angle is less than $\sim 20^\circ$ (ref. 15). Moving at nominally 500 km s^{-1} , it would take a spherically symmetric disturbance in the solar wind 17 h to travel from Cassini to Saturn.

Figure 1 shows the solar wind magnetic field, ion speed and

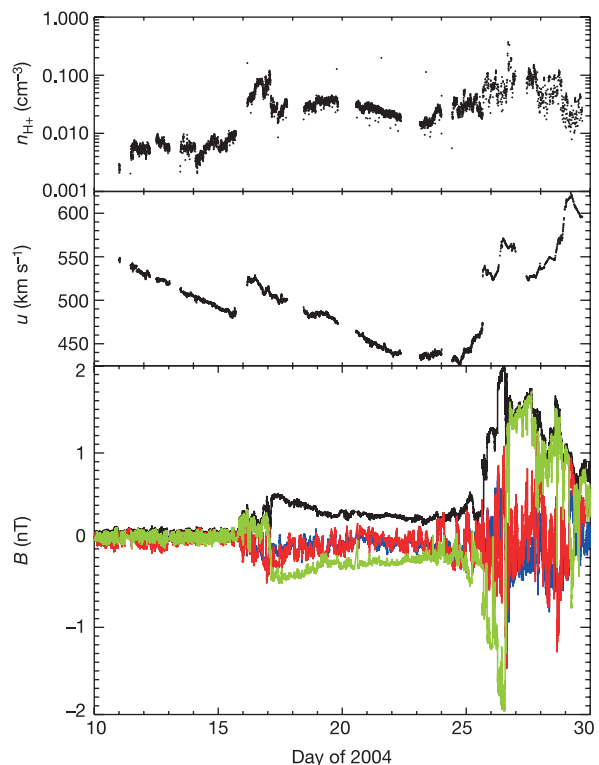


Figure 1 Solar wind conditions measured by Cassini. The top panel shows proton density, the middle panel the flow speed, and the lower panel the magnetic field. The black line shows the magnitude of the magnetic field. B_x is shown in blue, B_y in green and B_z in red. The magnetic field is in planetary coordinates, a right-handed system with z parallel to the axis of rotation (and magnetic moment), and the Sun in the $+x/z$ half-plane.

proton density. The solar wind speed is characterized by periods of gradual decrease followed by shocks on 15 and 25 January. Enhanced ion densities are associated with these shocks, as are enhancements in energetic particle flux and low frequency plasma waves. Energetic particle, wave and magnetic field data indicate another shock on 1 January, when the spacecraft orientation did not allow direct measurements of core solar wind protons. Following the 25 January shock, the plasma density remained high and variable until the end of the month, the magnetic field strength was enhanced and several subsequent shocks were encountered. These solar wind disturbances are consistent with corotating interaction regions. A bimodal solar wind structure as observed here is expected during the declining phase of the solar cycle magnetic sector boundaries, with a reversal of the dominant B_y component that occurred on 17 and 26 January. (B_y indicates the y component of the magnetic field; see Fig. 1 legend for coordinate system.) During the month of observations, the north–south component of the interplanetary magnetic field was weak, and the field was primarily in the east–west direction.

We have used Cassini measurements to estimate conditions at Saturn with a one-dimensional, numerical magneto-hydrodynamic model¹⁶. This model assumes spherical symmetry and propagates the measurements outward. It accounts for the fact that shocks propagate faster than the ambient flow speed, as well as for any evolution of the solar wind during the final 0.25 AU of its trip to Saturn. As the model assumes spherical symmetry, the three-dimensional structure of the solar wind may cause errors in the modelled propagation time. We expect that our estimated arrival times may be systematically late by up to roughly 10 h, but cannot be more exact owing to the unconstrained, three-dimensional structure of the solar wind.

According to this model, the shocks observed on 15 and 25 January reached Saturn at 12:45 UTC, 16 January, and 08:00 UTC,

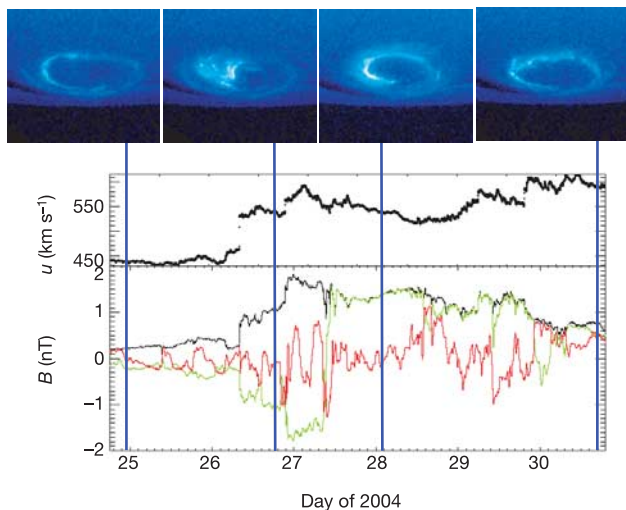


Figure 2 Comparison between HST images and solar wind conditions propagated to Saturn for the period 25–30 January 2004. For details of the HST images, see ref. 3. Thick blue vertical lines indicate the times of the four HST images shown above. The solar wind conditions are presented in the same format as Fig. 1, but density and the B_x component are omitted for clarity. In one-dimensional, time-dependent magneto-hydrodynamics with a uniform upstream boundary condition, the x component cannot be accurately modelled. The modelled propagation times may be systematically long by as much as 10 h. Non-radial orientation of the solar wind structures, typical of corotating interaction regions, could cause them to reach Saturn roughly five hours earlier than spherically symmetric calculations would suggest. Solar rotation and the 3.3° difference between the spacecraft's and Saturn's heliocentric longitude could also move arrival times earlier by approximately 6 h. Overall, the estimated arrival times may be systematically late by up to roughly 10 h.

26 January. Figure 2 shows the propagated plasma speed and the magnetic field, along with four of the HST images. The first image, 32 h before the arrival of the 25 January shock, is typical of other images taken during undisturbed solar wind conditions. The second image, 11 h after the shock arrival (+11 h), shows a brightened and partially filled auroral oval. In the third image, at +42 h, the oval has shifted to higher latitudes while remaining atypically bright. In the final image, at +107 h, the aurora has returned to a more typical state. The response to the 15 January shock appears similar, although both the shock and the auroral response were weaker, and the first image obtained after the shock was at +40 h (ref. 3).

Figure 3 shows a plot of propagated solar wind dynamic pressure against the input auroral power, estimated from the total ultraviolet brightness³. There is a very good correlation between auroral power and both the solar wind dynamic pressure, ρu^2 ($r = 0.93$), and the convection electric field, $\mathbf{u} \times \mathbf{B}$ ($r = 0.97$). (Here ρ is the mass density, u the flow velocity, r the linear correlation coefficient, $\mathbf{u}$ the flow velocity vector, and $\mathbf{B}$ the magnetic field vector.) These high correlation coefficients are, in part, a statistical artefact of the 25 January event, which produced the two unusually high-power points. However, this event is not the sole source of the correlations. Using only those observations with an auroral power below 5×10^{10} W (that is, using only the lower-power half of the measurements and thus excluding the unusual events), we obtain correlation coefficients of 0.62 and 0.92, respectively.

Correlations with auroral power also show a lack of control by the direction of the interplanetary magnetic field. In the case of Earth's magnetosphere, the dynamics correlate well with a gated solar wind electric field, $\mathbf{u} \times \mathbf{B} \cos^4(\theta/2)$, where θ is the projected angle between the interplanetary magnetic field and the planet's magnetic moment¹⁷. The θ term accounts for efficient, low-latitude reconnection when the interplanetary magnetic field is parallel to the planet's magnetic moment. For the January measurements, the correlation coefficient with this parameter is 0.74. This is a significantly lower value than that obtained without the θ term, suggesting that the orientation of the interplanetary magnetic field is not a controlling factor.

At the same time, the interplanetary magnetic field during January was primarily in the y direction, while the north–south component was weak and varying, with $\theta = 91 \pm 32^\circ$. An alternative explanation is that the $\cos^4$ gating function is not well-suited for the interaction between the solar wind and Saturn's magnetosphere. This relation was derived for the Earth and for a wide variety of solar wind magnetic field orientations, including the predominantly southward conditions that produce a strong magnetospheric

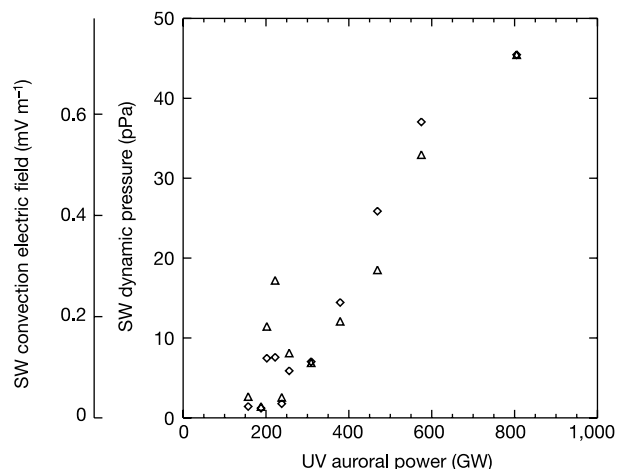


Figure 3 Correlation between HST-derived auroral input power and solar wind conditions. Triangles show the solar wind (SW) dynamic pressure, ρu^2 ; diamonds show the solar wind convection electric field, $|\mathbf{u} \times \mathbf{B}|$.

response. It is possible that the most appropriate gating function for values of θ near 90° differs from the $\cos^4$ relation best suited for the full range of θ .

It is interesting to note that the January observations occurred near southern summer, and that the average interplanetary magnetic field is almost purely azimuthal at Saturn's distance from the Sun. As a result, the dominant component of the field was perpendicular to Saturn's magnetic moment. Solar wind control of Saturn's magnetosphere could be significantly different near the equinoxes, when Saturn's axial tilt results in a 64.5° angle between the dominant, azimuthal field component and Saturn's magnetic moment. This would be a manifestation of a well-documented phenomenon at Earth¹⁸, but made more extreme owing to the alignment between Saturn's magnetic moment and spin axis.

Overall, the January Cassini and HST measurements indicate that Saturn's magnetosphere, although similar in some respects to those of the Earth and Jupiter, is not simply an intermediate case. It resembles Earth's magnetosphere in that its auroral dynamics are strongly driven by solar wind conditions. But it differs from Earth's interaction with the solar wind by virtue of an evidently weak influence of the direction of the interplanetary magnetic field. This may either be a characteristic of Saturn's magnetosphere, or be a result of differences between the solar wind at 10 AU and 1 AU (specifically, the weaker north-south component of the solar wind magnetic field). The latter possibility suggests that Saturn's magnetospheric response to the solar wind may be primarily driven by solar wind shocks, a process that is observed at the Earth² but generally overshadowed by the strong response of Earth's magnetosphere to the orientation of the interplanetary magnetic field. This apparent insensitivity to the magnetic field orientation and sensitivity to solar wind dynamic pressure and/or motional electric field make Saturn's magnetosphere resemble that of Jupiter more than that of Earth. In contrast to Jupiter, where the influence of the solar wind is weak and the aurorae are largely due to internal processes, the solar wind plays a controlling role in Saturn's auroral dynamics. □

Received 9 September; accepted 27 December 2004; doi:10.1038/nature03333.

1. Dungey, J. W. Interplanetary magnetic field and the auroral zone. *Phys. Rev. Lett.* **6**, 47–48 (1961).
2. Hill, T. W., Dessler, A. J. & Goertz, C. K. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) 353–394 (Cambridge Univ. Press, Cambridge, 1983).
3. Clarke, J. T. *et al.* Morphological differences of Saturn's ultraviolet aurorae and those of Earth and Jupiter. *Nature* doi:10.1038/nature03331 (this issue).
4. Meurant, M., Gerard, J.-C., Blockx, C., Hubert, B. & Coumans, V. Propagation of electron and proton shock-induced aurora and the role of the interplanetary magnetic field and solar wind. *J. Geophys. Res.* **109**, A10210, doi:10.1029/2004JA010453 (2004).
5. Cowley, S. W. H. & Bunce, E. J. Origin of the main auroral oval in Jupiter's coupled magnetosphere-ionosphere system. *Planet. Space Sci.* **49**, 1067–1088 (2001).
6. Zarka, P. & Genova, F. Low frequency jovian emissions and solar wind magnetic sector structure. *Nature* **306**, 767–768 (1983).
7. Desch, M. D. & Barrow, C. H. Direct evidence for solar wind control of Jupiter's hectometric wavelength radio emission. *J. Geophys. Res.* **89**, 6819–6823 (1984).
8. Gurnett, D. A. *et al.* Control of Jupiter's radio emissions and aurorae by the solar wind. *Nature* **415**, 985–987 (2002).
9. Kivelson, M. G. & Southwood, D. J. First evidence of IMF control of Jovian magnetospheric boundary locations: Cassini and Galileo magnetic field measurements compared. *Planet. Space Sci.* **51**, 891–898 (2003).
10. Southwood, D. J. & Kivelson, M. G. A new perspective concerning the influence of the solar wind on the Jovian magnetosphere. *J. Geophys. Res.* **106**, 6123–6130 (2001).
11. Cowley, S. W. H. & Bunce, E. J. Modulation of Jupiter's main auroral oval emissions by solar wind induced expansions and compressions of the magnetosphere. *Planet. Space Sci.* **51**, 57–79 (2003).
12. Desch, M. D. Evidence for solar wind control of Saturn radio emission. *J. Geophys. Res.* **87**, 4549–4554 (1982).
13. Young, D. T. *et al.* Cassini plasma spectrometer investigation. *Space Sci. Rev.* (in the press).
14. Kurth, W. S. *et al.* An Earth-like correspondence between Saturn's auroral features and radio emission. *Nature* doi:10.1038/nature03334 (this issue).
15. Hanlon, P. G. *et al.* On the evolution of the solar wind between 1 and 5 AU at the time of the Cassini-Jupiter flyby: multi spacecraft observations of ICMEs including the formation of a merged interaction region. *J. Geophys. Res.* **109**, A09S03, doi:10.1029/2003JA010112 (2004).
16. Tóth, G. General code for modeling MHD flows on parallel computers: Versatile advection code. *Astrophys. Lett. Commun.* **34**, 245–250 (1996).
17. Akasofu, S.-I. The solar wind-magnetosphere energy coupling and magnetospheric disturbances. *Planet. Space Sci.* **28**, 495–509 (1980).
18. Russell, C. T. & McPherron, R. L. Semiannual variation of geomagnetic activity. *J. Geophys. Res.* **78**, 92–108 (1973).

Acknowledgements F.J.C., B.L.B., J.T.S. and D.T.Y. were supported by NASA through a Jet Propulsion Laboratory contract with SWRI; D.G.M., K.C.H. and W.S.K. were supported through other NASA/JPL contracts; P.G.H. was supported by a PPARC-UK quota studentship; and J.C.G. and D.G. were supported by the Belgian Foundation for Scientific Research (FNRS) and the PRODEX program of the ESA. This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by the AURA, Inc., for NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.J.C. (fcrary@swri.edu).

An Earth-like correspondence between Saturn's auroral features and radio emission

W. S. Kurth¹, D. A. Gurnett¹, J. T. Clarke², P. Zarka³, M. D. Desch⁴, M. L. Kaiser⁴, B. Cecconi¹, A. Lecacheux⁵, W. M. Farrell⁴, P. Galopeau⁵, J.-C. Gérard⁶, D. Grodent⁶, R. Prangé³, M. K. Dougherty⁷ & F. J. Crary⁸

¹Department of Physics and Astronomy, The University of Iowa, Iowa City, Iowa 52242, USA

²Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

³Space Research Department, Observatoire de Paris, 92195 Meudon, France

⁴NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

⁵CETP/UVSQ, 78140 Velizy, France

⁶LPAP, Université de Liège, allée du 6 août, 17, B-4000 - Liège, Belgium

⁷Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

⁸Southwest Research Institute, Culebra Road, San Antonio, Texas 78288, USA

Saturn is a source of intense kilometre-wavelength radio emissions that are believed to be associated with its polar aurorae^{1,2}, and which provide an important remote diagnostic of its magnetospheric activity. Previous observations implied that the radio emission originated in the polar regions, and indicated a strong correlation with solar wind dynamic pressure^{1,3–7}. The radio source also appeared to be fixed near local noon and at the latitude of the ultraviolet aurora^{1,2}. There have, however, been no observations relating the radio emissions to detailed auroral structures. Here we report measurements of the radio emissions, which, along with high-resolution images of Saturn's ultraviolet auroral emissions⁸, suggest that although there are differences in the global morphology of the aurorae, Saturn's radio emissions exhibit an Earth-like correspondence between bright auroral features and the radio emissions. This demonstrates the universality of the mechanism that results in emissions near the electron cyclotron frequency narrowly beamed at large angles to the magnetic field^{9,10}.

The first studies of Saturn's primary radio emission (Saturn kilometric radiation, SKR) were by the Voyager spacecraft in the early 1980s^{1,3}. This emission had some features that were more like those of Earth than of Jupiter: the SKR source appears to be fixed in local time, favouring a region near local late morning or noon, but varying with source latitude². It is also strongly correlated with solar wind parameters, such as the dynamic pressure⁷. However, the SKR also has some features that are more like those of Jupiter than of Earth: it was found to be strongly modulated by planetary rotation, despite the fact that models of Saturn's magnetic field are axisymmetric¹¹. This has suggested an as-yet-unobserved 'active sector' or possible 'magnetic anomaly' rotating with the planet^{12,13}.

The measurements shown here are primarily from the radio

and plasma wave science instrument on Cassini¹⁴ and the Space Telescope Imaging Spectrograph (STIS) on the Hubble Space Telescope (HST)⁸. Cassini continuously monitors the intensity and polarization of SKR (typically observed between ~ 20 and 800 kHz). During the month of January 2004, the observed SKR was almost entirely left-hand circularly polarized, consistent with emissions from Saturn's southern hemisphere. This is largely due to the fact that Cassini was at a saturnian latitude of about -16° . We use this dominance of left-hand circularly polarized emission to derive the total hemispherical integrated radio flux at Cassini by using only this polarization to eliminate contributions from solar type III bursts and a few other low-frequency emissions. Similarly, the ultraviolet observations are restricted to the southern hemisphere by the viewing geometry from Earth. In Fig. 1 we compare the integrated power per unit solid angle with the power input to the aurora. The plot in Fig. 1a shows a general correspondence between the auroral input power and SKR. Given some Voyager estimates of the SKR rotationally averaged solid angle of 2π sr (ref. 6), we obtain emission efficiencies of about 0.5%. This is comparable to the efficiency of generating terrestrial kilometric radiation¹⁵ ($\sim 1\%$).

The plot of the auroral input power and SKR emitted power,

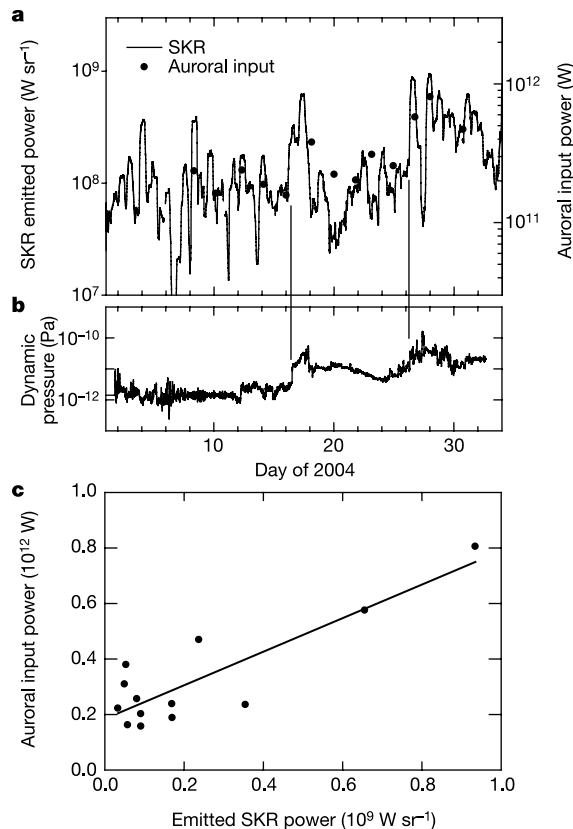


Figure 1 Correlations between auroral input power and emitted SKR power. **a**, Time-averaged power emitted as SKR as a function of time, overplotted with the estimated auroral input power based on the ultraviolet aurora integrated intensity. The radio power is a sliding 10 h 40 min (Saturn's approximate rotational period) average, using the left-hand circularly polarized component of the wave spectrum integrated between 10 kHz and 1 MHz. The rotational average is used because Voyager observations indicated that for short averaging intervals of the SKR power (less than several hours) there is no correlation between intensities observed by two widely spaced observers⁵. The auroral input power is estimated from the integrated ultraviolet intensity over the auroral region assuming electron precipitation⁸. **b**, Solar wind dynamic pressure determined from Cassini measurements of the solar wind density and speed, and propagated to Saturn with a simple radial magnetohydrodynamic model¹⁶. **c**, Plot comparing the auroral input power and emitted radio power.

shown in Fig. 1c, provides a more quantitative analysis of the relation between these two parameters. The linear fit to these data gives a positive correlation coefficient, r , of 0.87. However, without the two large-amplitude points obtained on 26 and 28 January, the remaining points are not correlated. Perhaps one reason for this is the large disparity in the measurement duration for the two parameters; 10.67 h for SKR and about 20 min for the ultraviolet intensity. On the other hand, the two highest-amplitude points from the HST observations have the largest signal-to-noise ratio and, hence, are the most reliable. During the period encompassing the 26 and 28 January images, Saturn was under the influence of a corotating interaction region with high field amplitudes and increased solar wind dynamic pressure¹⁶ (Fig. 1b).

A five-orbit sequence of HST observations acquired on 8 January 2004 is useful in understanding in greater detail the relation between the ultraviolet emissions and, in this case, relatively weak SKR emissions. The first image (Fig. 2, image (1) in the top row) shows that even though there is a relatively bright aurora near dawn, there is very little corresponding radio emission (Fig. 2, bottom panel). The next three images, (2) to (4) in Fig. 2, show the bright spot moving to increasing local times and latitudes while the radio emission spectrum intensifies and expands to lower frequencies. The last image (5) shows a weaker auroral emission that has moved off to the local afternoon just near the end of the SKR event. The ultraviolet power during this series of observations varies by more than a factor of three.

As Cassini is nearly fixed in position with respect to Saturn during this time, we can explore the relationship between the bright auroral features, which appear to be rotating at about 30–75% of corotation⁸ with the planet, and the observed SKR, under the assumption that the SKR source, like terrestrial kilometric radiation, is tied to bright auroral features^{15,17,18}. Using the Z3 model of Saturn's magnetic field¹¹, we can determine the geometry of magnetic field lines connected to the centroids of three bright auroral 'spots' (A, B, C) seen in each of the five ultraviolet images in Fig. 2. The altitude of the source is where the emission frequency equals the electron cyclotron frequency, f , given by $f = 28|B|$, where

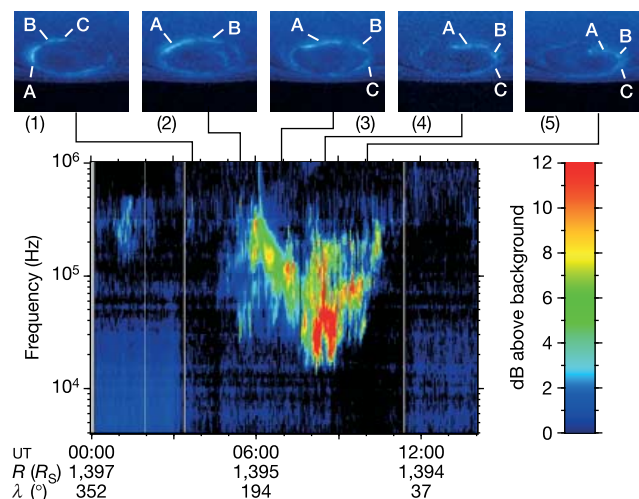


Figure 2 Detailed SKR observations from 8 January and the associated ultraviolet images. Top, five images covering about 70% of a Saturn rotation. Bottom, spectrogram displaying the intensity of radio waves (using the colour bar at the right) as a function of frequency (ordinate) and time (abscissa). At the bottom, R refers to the radial distance from Saturn in units of Saturn radii (R_S), and λ is longitude. The images⁸ are obtained by the STIS instrument on the Hubble Space Telescope with the image brightness scaled to the observed intensity of ultraviolet emissions. Lines relate each image to the respective time on the spectrogram. The letters A, B and C indicate the centroid of bright spots referred to in Fig. 3.

f is in Hz and B (the magnetic field strength) is in nT. Without anticipating the beaming properties of the emission, we can determine, from purely geometric considerations, the emission angle with respect to the magnetic field for the radio source to be visible from Cassini at each frequency and for each of the images. The resulting acceptable combinations of beaming angles and frequencies are shown as thin lines in Fig. 3 for each image (1 to 5). Panels a, b and c show the respective results for spots labelled A, B and C in Fig. 2. The observed frequency ranges are plotted with thick lines. Beaming angles less than 90° in the southern hemisphere imply propagation towards the planet and increasing field strengths. As the extraordinary mode cannot propagate where $f < f_{ce}$ (f_{ce} is the electron cyclotron frequency), these solutions are unphysical. A source on field lines threading spots B or C cannot account for all of the observed emission observed in the five images; but a source

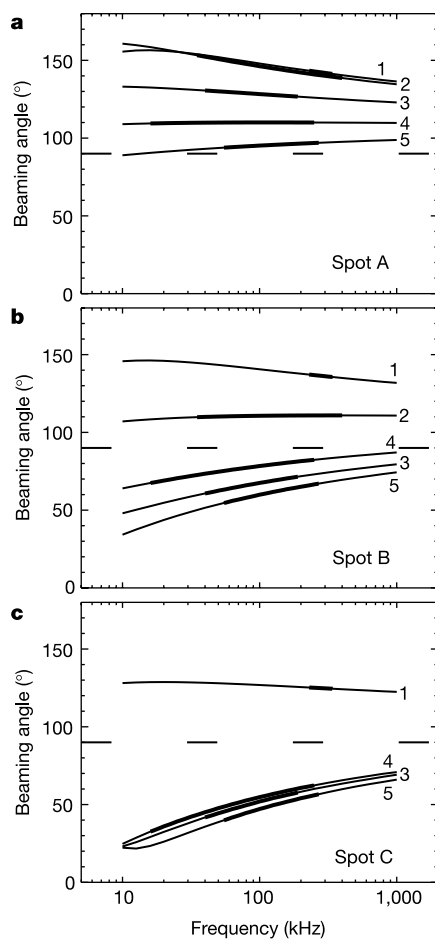


Figure 3 Beaming angles required to illuminate Cassini from radio sources associated with bright spots identified in Fig. 2. **a**, For the bright spot identified with an A in the images in Fig. 2; the thin lines indicate the beaming angle required for a radio wave to propagate from a source threading the centre of the bright spot labelled A to Cassini as a function of frequency. This construction is made assuming the emission frequency is at the electron cyclotron frequency and using the Z3 magnetic field model for Saturn¹¹. The thick lines represent the range of frequencies actually observed at Cassini. The numbers refer to the five images at the top of Fig. 2. **b**, Similar to **a** but the geometry is for sources on a field line threading the bright spot labelled B in Fig. 2. **c**, Similar to **a** and **b** but for the bright spot identified with a C. The identification of this particular feature from one image to the next is less clear than for the others. Note that beaming angles greater than 90° (indicated by the dashed line) are directed away from Saturn and away from stronger field regions. Beaming angles less than 90° are directed downward towards the planet in order to intersect Cassini. The downward propagation, however, is forbidden by the presence of Saturn and the stronger magnetic fields near it; hence, these waves cannot reach Cassini.

associated with bright spot A can. Carrying this analysis further, it is possible to consider all the possible sources connected to spot A, for each image. We find that all of the observed emission can be explained by sources on field lines in the vicinity of spot A using only a narrow range of beaming angles, 110 – 125° . Hence it is possible to associate SKR with bright aurorae assuming relatively narrow emission cones.

Cassini does not observe the radio emissions for the entire frequency range all of the time. This is because the cyclotron maser instability produces many fluctuating, localized sources emitting in very narrow beams; localized and probably time variable differences between the Z3 model and the actual field also affect the visibility of the emissions. It is clear that the correlation between the ultraviolet and SKR power is limited, as HST can observe the entire auroral oval but the radio emissions are not always beamed in the direction of Cassini. Nevertheless, Fig. 3 demonstrates that SKR can be tied to bright aurorae, as is the case at Earth^{15,17,18}. The thick lines in Fig. 3 for angles $>90^\circ$ may be interpreted as an ensemble average of emission extending in altitude from field lines threading the vicinity of the bright auroral spot A to account for the range in frequencies observed.

As spot A rotates through a limited range of local times during the sequence of images obtained on 8 January⁸, and the local time range is generally consistent with that obtained for the SKR source region from Voyager observations, we now have sufficient information to clarify one of the conclusions based on Voyager observations—that the source is fixed in local time. We can now say that the source could, indeed, move as the bright spots are seen to move, but within a limited range of local times.

There is additional evidence that beaming is important for the detection of SKR. The large dip in SKR power seen on day 27 in Fig. 1a corresponds to a ‘missing’ SKR event; the usual modulation of SKR seems to skip a rotation on 27 January. However, the Unified Radio and Plasma Wave experiment¹⁹ on the Ulysses²⁰ spacecraft, which is at slightly lower saturnian latitudes than Cassini, does see SKR during the expected time frame. During this time period, Saturn is under the influence of higher solar wind dynamic pressure associated with a corotating interaction region in the solar wind¹⁶. We believe that the higher solar wind dynamic pressure at this time distorts the field and moves the SKR beam away from Cassini. This evidence, although anecdotal, confirms that beaming must be considered in any detailed comparison of auroral brightness and SKR.

The bright aurora observed on 26 and 28 January⁸ shows poleward expansion and is accompanied by intense SKR that spreads to lower frequencies. The total SKR power is a function of bandwidth, with larger fluxes seen with broader bandwidths. Clarke *et al.*⁸ show that the auroral latitude and auroral power are correlated. At Earth, expansion of kilometric radiation to low frequencies is correlated with higher fluxes of auroral bremsstrahlung X-rays²¹, so the association between the increase of the SKR bandwidth and intense aurora is a similarity between Earth and Saturn. Such a correlation between increased precipitation and expansion of SKR to lower frequencies is probably a result of the Knight relationship²²; increased auroral currents require an increase in the upper altitude limit of the auroral acceleration region²³. Consequently, the SKR source region probably expands to higher altitudes, where f_{ce} is lower. It is also likely that beaming plays a role in this frequency variation, given the poleward expansion of the aurora and that significant distortions in the magnetic field are possible as interplanetary shocks, varying interplanetary field orientations and solar wind dynamic pressure variations interact with the magnetosphere. □

Received 13 August; accepted 4 December 2004; doi:10.1038/nature03334.

1. Kaiser, M. L. *et al.* in *Saturn* (eds Gehrels, T. & Matthews, M. S.) 378–415 (Univ. Arizona Press, Tucson, 1984).

2. Galopeau, P. H. M., Zarka, P. & Le Quéau, D. Source location of Saturn's kilometric radiation: The Kelvin-Helmholtz instability hypothesis. *J. Geophys. Res.* **100**, 26397–26410 (1995).
3. Warwick, J. W. *et al.* Planetary radio astronomy observations from Voyager 1 near Saturn. *Science* **212**, 239–243 (1981).
4. Galopeau, P. H. M. & Lecacheux, A. Variations in Saturn's radio rotation period measured at kilometer wavelengths. *J. Geophys. Res.* **105**, 13089–13101 (2000).
5. Zarka, P. in *Planetary Radio Emissions II* (eds Rucker, H. O., Bauer, S. J. & Pedersen, B. M.) 327–342 (Austrian Acad. Sci. Press, Vienna, 1988).
6. Zarka, P. Auroral radio emissions at the outer planets: Observations and theories. *J. Geophys. Res.* **103**, 20159–20194 (1998).
7. Desch, M. D. Evidence for solar wind control of Saturn radio emission. *J. Geophys. Res.* **87**, 4549–4554 (1982).
8. Clarke, J. T. *et al.* Morphological differences of Saturn's ultraviolet aurorae and those of Earth and Jupiter. *Nature* doi:10.1038/nature03331 (this issue).
9. Wu, C. S. & Lee, L. C. A theory of terrestrial kilometric radiation. *Astrophys. J.* **230**, 621–626 (1979).
10. Galopeau, P., Zarka, P. & Le Quéau, D. Theoretical model of Saturn's kilometric radiation spectrum. *J. Geophys. Res.* **94**, 8739–8755 (1989).
11. Connerney, J. E. P., Ness, N. F. & Acuña, M. H. Zonal harmonic model of Saturn's magnetic field from Voyager 1 and 2 observations. *Nature* **298**, 44–46 (1982).
12. Galopeau, P., Ortega-Molina, A. & Zarka, P. Evidence of Saturn's magnetic field anomaly from SKR high-frequency limit. *J. Geophys. Res.* **96**, 14129–14140 (1991).
13. Galopeau, P. & Zarka, P. Reply to the comment by J. E. P. Connerney and M. D. Desch on "Evidence of Saturn's magnetic field anomaly from SKR high-frequency limit". *J. Geophys. Res.* **97**, 12291–12297 (1992).
14. Gurnett, D. A. *et al.* The Cassini radio and plasma wave science investigation. *Space Sci. Rev.* (in the press).
15. Gurnett, D. A. The Earth as a radio source: Terrestrial kilometric radiation. *J. Geophys. Res.* **79**, 4227–4238 (1974).
16. Cray, F. J. *et al.* Solar wind dynamic pressure and electric field as the main factors controlling Saturn's aurorae. *Nature* doi:10.1038/nature03333 (this issue).
17. Huff, R. L., Calvert, W., Craven, J. D., Frank, L. A. & Gurnett, D. A. Mapping of auroral kilometric radiation sources to the aurora. *J. Geophys. Res.* **93**, 11445–11454 (1988).
18. de Feraudy, H., Bahnsen, A. & Jespersen, M. in *Planetary Radio Emissions II* (eds Rucker, H. O., Bauer, S. J. & Pedersen, B. M.) 41–60 (Austrian Acad. Sci. Press, Vienna, 1988).
19. Stone, R. G. *et al.* The unified radio and plasma wave investigation. *Astron. Astrophys. Suppl. Ser.* **92**, 291–316 (1992).
20. Wenzel, K. P., Marsden, R. G., Page, D. E. & Smith, E. J. The Ulysses mission. *Astron. Astrophys. Suppl. Ser.* **92**, 207–219 (1992).
21. Imhof, W. L. *et al.* The dependence of AKR production on the intensity and energy spectra of auroral bremsstrahlung. *J. Geophys. Res.* **108**, doi:10.1029/2002JA009274 (2003).
22. Knight, S. Parallel electric fields. *Planet. Space Sci.* **21**, 741–750 (1973).
23. Cowley, S. W. H. & Bunce, E. J. Origin of the main auroral oval in Jupiter's coupled magnetosphere-ionosphere system. *Planet. Space Sci.* **49**, 1067–1088 (2001).

Acknowledgements This research was supported by NASA through contracts with the Jet Propulsion Laboratory. J.-C.G. and D.G. are supported by the Belgian Fund for Scientific Research (FNRS) and partly funded by the PRODEX programme of the European Space Agency. This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by AURA, Inc., for NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.S.K. (william-kurth@uiowa.edu).

A continuous-wave Raman silicon laser

Haisheng Rong¹, Richard Jones¹, Ansheng Liu¹, Oded Cohen², Dani Hak², Alexander Fang¹ & Mario Paniccia¹

¹Intel Corporation, 2200 Mission College Blvd, CHP3-109, Santa Clara, California 95054, USA

²Intel Corporation, SBI Park Har Hotzvim, Jerusalem, 91031, Israel

Achieving optical gain and/or lasing in silicon has been one of the most challenging goals in silicon-based photonics^{1–3} because bulk silicon is an indirect bandgap semiconductor and therefore has a very low light emission efficiency. Recently, stimulated Raman scattering has been used to demonstrate light amplification and lasing in silicon^{4–9}. However, because of the nonlinear optical loss associated with two-photon absorption (TPA)-induced free

carrier absorption (FCA)^{10–12}, until now lasing has been limited to pulsed operation^{8,9}. Here we demonstrate a continuous-wave silicon Raman laser. Specifically, we show that TPA-induced FCA in silicon can be significantly reduced by introducing a reverse-biased p-i-n diode embedded in a silicon waveguide. The laser cavity is formed by coating the facets of the silicon waveguide with multilayer dielectric films. We have demonstrated stable single mode laser output with side-mode suppression of over 55 dB and linewidth of less than 80 MHz. The lasing threshold depends on the p-i-n reverse bias voltage and the laser wavelength can be tuned by adjusting the wavelength of the pump laser. The demonstration of a continuous-wave silicon laser represents a significant milestone for silicon-based optoelectronic devices.

The continuous-wave (c.w.) silicon Raman laser is constructed from a low-loss silicon-on-insulator (SOI) rib waveguide whose facets are coated with multilayer dielectric films. The front facet coating is dichroic, having a reflectivity (R_f) of ~71% for the Raman/Stokes wavelength of 1,686 nm and ~24% for the pump wavelength of 1,550 nm. The back facet has a broadband high-reflectivity coating (R_b) of ~90% for both pump and Raman wavelengths (Fig. 1a). These waveguide facet reflectivities were determined using a Fabry-Pérot resonance technique².

The silicon rib waveguide is fabricated on the (100) surface of an undoped SOI substrate using standard photolithographic patterning and reactive ion etching techniques. We designed the waveguide dimensions with the goal of obtaining a small cross-section for minimizing the required optical power to achieve the lasing threshold, but not so small as to cause high transmission loss. A cross-section scanning electron microscope image of a typical p-i-n waveguide is shown in Fig. 1b. The rib waveguide dimensions are: rib width (W) ~1.5 μm ; height (H) ~1.55 μm ; and etch depth (h) ~0.7 μm . The effective core area¹³ of the waveguide is calculated to

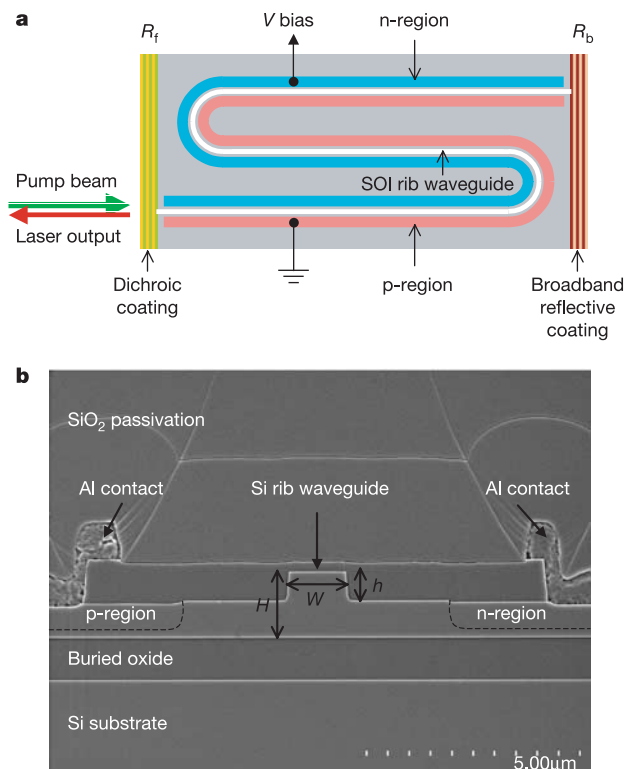


Figure 1 Silicon waveguide used in the Raman laser experiment. **a**, Schematic layout of the silicon waveguide laser cavity with optical coatings applied to the facets and a p-i-n structure along the waveguide. **b**, Scanning electron microscope cross-section image of a silicon rib waveguide with a p-i-n diode structure.

be $\sim 1.6 \mu\text{m}^2$. The waveguide was formed in an S-shaped curve with a total length of 4.8 cm and a bend radius of $400 \mu\text{m}$ (Fig. 1a). The straight sections of the waveguide are oriented along the [011] crystallographic direction.

A p-i-n diode structure was designed to reduce the nonlinear optical loss due to the TPA-induced FCA. The p-i-n structure was formed by implanting boron and phosphorus in the slab on either side of the rib waveguide (Fig. 1a, b) with a doping concentration of $\sim 1 \times 10^{20} \text{cm}^{-3}$. The separation between the p- and n-doped regions was designed to be $\sim 6 \mu\text{m}$. Ohmic contacts were formed by depositing aluminium films on the surface of the p- and n-doped regions. This was followed by a SiO_2 passivation layer deposition. The doped regions and the metal contacts at the designed separation had negligible effect on the propagation loss of the waveguide because the optical mode is tightly confined in the waveguide. This was verified experimentally. The linear optical transmission loss of the S-bend waveguide was measured to be 0.35 dB cm^{-1} using the Fabry–Pérot resonance technique².

When a reverse bias voltage is applied to the p-i-n diode, the TPA-generated electron–hole pairs can be swept out of the silicon waveguide by the electric field between the p- and n-doped regions. Thus the effective carrier lifetime, representing the lifetime of the free carrier's interaction with the optical mode in the waveguide region, reduces with increased bias voltage. This has been experimentally verified by comparing the measured nonlinear transmission of a silicon waveguide with modelling^{9,14}. The measured photocurrent in the reverse-biased p-i-n diode scales with the square of the light power inside the waveguide, indicating that the charge carriers are generated by the TPA process¹⁰. At a reverse-bias voltage of 25 V, the effective carrier lifetime is reduced to $\sim 1 \text{ ns}$ compared to the free carrier lifetime of several tens of nanoseconds in ordinary silicon rib waveguides^{10,11}. Before performing the lasing experiment, the single-pass c.w. Raman gain of a p-i-n silicon waveguide was measured in a pump–probe experiment¹⁴, showing a single-pass net gain of $>3 \text{ dB}$ at a reverse-bias voltage of 25 V and a pump power of $\sim 700 \text{ mW}$ coupled into the waveguide.

Figure 2 is a schematic of the Raman laser experiment. A c.w. external cavity diode laser (ECDL) at 1,550 nm is amplified by an erbium-doped fibre amplifier system to produce a pump beam of up to 3 W. The pump beam passes through a polarization controller followed by a thin-film-based wavelength de-multiplexer and is coupled into the waveguide cavity by a lensed fibre through the dichroic-coated front facet. The Raman laser output and the reflected pump beam are coupled back into the lensed fibre, and separated through the wavelength de-multiplexer. The extracted laser output from the reflection port of the de-multiplexer is further filtered by a long-wavelength pass filter before being detected by a power meter or optical spectrum analyser. The coupling loss

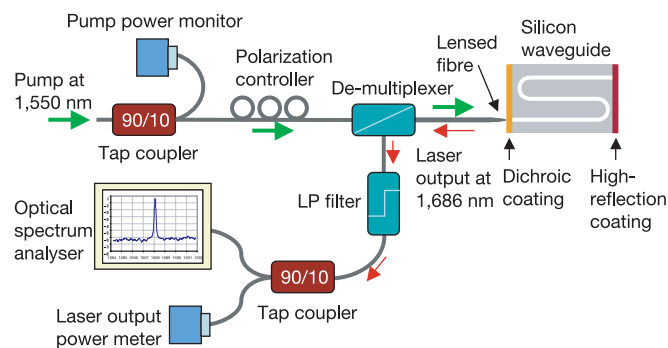


Figure 2 Schematic set-up of the silicon Raman laser experiment. See main text for details.

between the lensed fibre and the waveguide was measured to be $\sim 4 \text{ dB}$ and the insertion loss of the de-multiplexer and the long-wavelength pass filter is $\sim 0.6 \text{ dB}$. The silicon chip is mounted on a thermo electric cooler and kept at a constant temperature of 25°C .

At the pump wavelength, a low finesse cavity is formed by the low reflectivity front facet and high reflectivity back facet. This configuration allows the cavity enhancement effect^{15–17} of the pump power to be used to lower the lasing threshold. When the pump laser is tuned to the resonance of the cavity, the circulating power inside the waveguide cavity is enhanced, and the effective mean internal power (I_{eff}) inside the cavity can be expressed as¹⁸:

$$I_{\text{eff}} = I_i \frac{1 - e^{-\alpha L} (1 - R_f)(1 + R_b e^{-\alpha L})}{\alpha L (1 - \sqrt{R_f R_b} e^{-\alpha L})^2}$$

where I_i is the incident pump power (less coupling loss), R_f and R_b are the reflectivity of the front and back facet respectively, α is the absorption coefficient, and L is the waveguide length. At low power, we can estimate the power enhancement factor $M = I_{\text{eff}}/I_i$ to be ~ 2.2 , using our experimental parameters. At high power, however, the absorption coefficient α increases owing to the TPA-induced nonlinear absorption, and M reduces accordingly.

Figure 3 plots the Raman laser output power versus the input pump power (I_i) coupled into the laser cavity depicted in Fig. 1a, at two different reverse-bias voltages applied to the p-i-n diode. In the experiment, the pump beam polarization is adjusted with a polarization controller and its wavelength is fine-tuned to the cavity resonance to take advantage of the cavity enhancement of the pump and maximize the laser output. The Raman laser frequency is 15.6 THz lower than that of the pump laser. We see from Fig. 3 that the lasing threshold reduces with increasing reverse-bias voltage. The lasing threshold is $\sim 280 \text{ mW}$ with a 5-V bias and $\sim 180 \text{ mW}$ with a 25-V bias. The lower threshold and higher laser output power with higher reverse-bias voltage are expected because the effective carrier lifetime is shorter, resulting in lower nonlinear loss and higher gain¹⁴. For this waveguide cavity, the total loss of the feedback mirrors ($R_f = 71\%$ and $R_b = 90\%$) at the lasing wavelength is $\sim 2 \text{ dB}$, so a $\sim 1\text{-dB}$ single-pass net gain is needed to reach lasing threshold. From previous measurements¹⁴, a $\sim 1\text{-dB}$ net c.w. gain is obtained at a pump power of $\sim 400 \text{ mW}$ with 25-V reverse bias and $\sim 600 \text{ mW}$ with 5-V reverse bias. Taking into account the cavity enhancement factor of ~ 2.2 , the lasing thresholds are

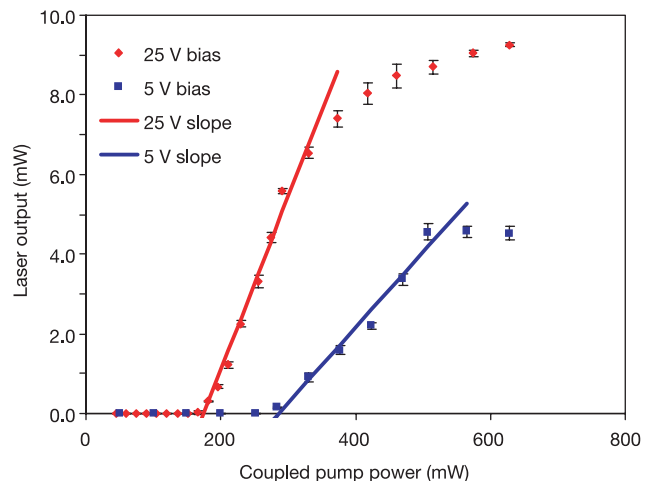


Figure 3 Silicon Raman laser output power as a function of the input pump power at a reverse bias of 25 and 5 V. The pump wavelength is 1,550 nm and the laser wavelength is 1,686 nm. The slope efficiency (single side output) is 4.3% for 25-V bias and 2% for 5-V bias. Error bars represent standard deviations.

expected to be ~ 182 and ~ 273 mW for reverse biases of 25 and 5 V, respectively, which is consistent with our measurements. The slope efficiency (single side output) above threshold is $\sim 4.3\%$ with a reverse bias of 25 V, and 2% with a reverse bias of 5 V. Figure 3 also shows that the laser output power begins to saturate at a pump power of >400 mW with a 25-V bias and at a pump power of >500 mW with a 5-V bias. This is primarily due to the nonlinear loss caused by the TPA-induced FCA as a consequence of the non-zero carrier lifetime of the p-i-n waveguide, reducing the net gain at higher pump powers. In addition, the cavity enhancement factor M for the pump reduces with increasing nonlinear loss, lowering the effective pump power in the cavity.

The spectrum of the laser output is measured with a confocal scanning Fabry-Pérot spectrum analyser with a free spectral range of 8 GHz and finesse of 100. Figure 4a shows the Raman laser spectrum measured at a pump power of ~ 400 mW and a reverse bias of 25 V. As shown, the laser has single-mode output, that is, no other cavity modes with expected mode spacing of 0.9 GHz for a 4.8-cm-long silicon waveguide cavity appear within the free spectral range (8 GHz) of the spectrum analyser. The measured linewidth of 80 MHz is limited by the resolution of the spectrum analyser. This was verified by recording the spectrum of a narrow-band laser of linewidth <1 MHz with the same spectrum analyser.

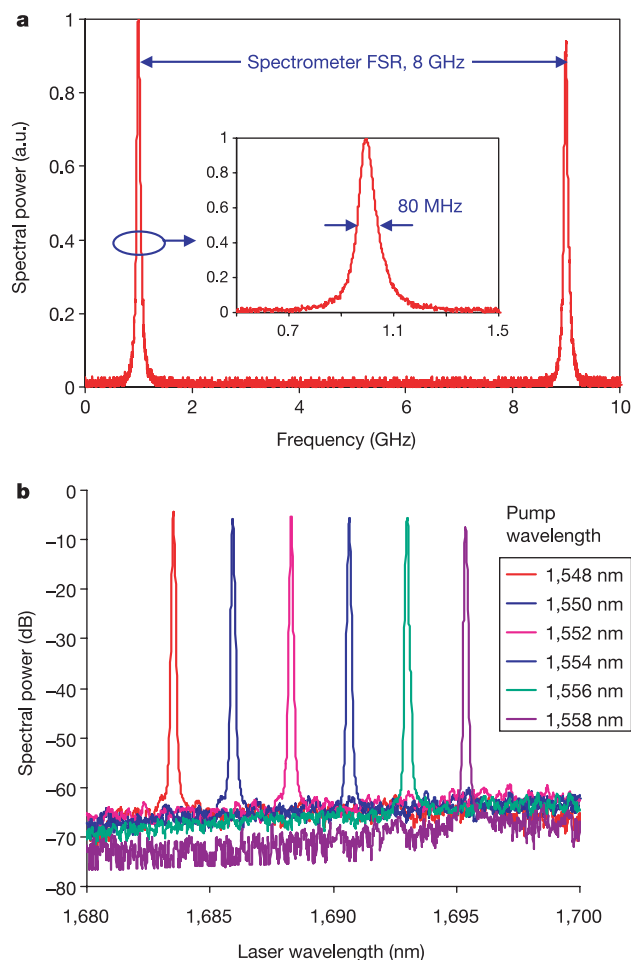


Figure 4 Silicon Raman laser spectra. **a**, Measured with a confocal scanning Fabry-Pérot spectrum analyser. Displayed is a scan over one free spectral range of 8 GHz showing single-mode operation of the laser. The measured linewidth of 80 MHz is limited by the resolution of the spectrum analyser. **b**, Measured with a grating-based optical spectrum analyser at different pump wavelengths from 1,548 nm to 1,558 nm in 2-nm steps, showing a side mode suppression of >55 dB.

Figure 4b is a plot of the laser output spectrum measured with a grating-based optical spectrum analyser with 0.07-nm resolution at various pump wavelengths. The spectra shown were obtained by changing the wavelength of the pump laser seed (ECDL) from 1,548 to 1,558 nm in 2-nm steps. The input pump power was ~ 400 mW and a reverse bias of 25 V was applied to the p-i-n diode. The Raman laser output has side-mode suppression of over 55 dB and its centre wavelength corresponds to the appropriate Stokes shift for each pump wavelength. The displayed linewidth is limited by the resolution of the spectrum analyser. The small fluctuation in output power is due to the wavelength dependence of the insertion loss of the de-multiplexer, the long-wavelength pass filter, and the gain of erbium-doped fibre amplifiers.

This first demonstration of c.w. Raman lasing in silicon represents a significant milestone towards producing fully integrated monolithic photonic chips. The performance of this silicon Raman laser could be further improved by optimizing cavity mirror and cavity length design. The threshold power could be reduced by using a waveguide with smaller cross-sectional dimensions and/or by introducing a larger cavity enhancement for the pump beam. The fibre to waveguide coupling efficiency could be improved by adding a mode converter in the waveguide¹⁹. In addition, with optimization of the p-i-n diode design, it may be possible to further reduce the effective carrier lifetime to below 1 ns. This would reduce the saturation effect and thus increase the laser output power. The multilayer coating approach used to form the cavity mirrors could also be replaced with waveguide Bragg reflectors^{20,21}, ring or microdisk resonator architectures^{22–25}; this could provide a platform for monolithic integration of silicon-based optoelectronics. □

Received 29 December 2004; accepted 10 January 2005; doi:10.1038/nature03346.

- Pavesi, L. & Gaponenko, S. (eds) *Towards the First Silicon Laser* (NATO Science Series, Kluwer, Dordrecht, 2003).
- Reed, G. T. & Knights, A. P. *Silicon Photonics: An Introduction* (John Wiley, Chichester, UK, 2004).
- Pavesi, L. & Lockwood, D. J. (eds) *Silicon Photonics* (Springer, Berlin, 2004).
- Liu, A., Rong, H., Paniccia, M., Cohen, O. & Hak, D. Net optical gain in a low loss silicon-on-insulator waveguide by stimulated Raman scattering. *Opt. Expr.* **12**, 4261–4267 (2004).
- Xu, Q., Almeida, V. & Lipson, M. Time-resolved study of Raman gain in highly confined silicon-on-insulator waveguides. *Opt. Expr.* **12**, 4437–4442 (2004).
- Liang, T. K. & Tsang, H. K. Efficient Raman amplification in silicon-on-insulator waveguides. *Appl. Phys. Lett.* **85**, 3343–3345 (2004).
- Boyras, O. & Jalali, B. Demonstration of 11 dB fiber-to-fiber gain in a silicon Raman amplifier. *IEEE Elect. Expr.* **1**, 429–434 (2004).
- Boyras, O. & Jalali, B. Demonstration of a silicon Raman laser. *Opt. Expr.* **12**, 5269–5273 (2004).
- Rong, H. *et al.* An all-silicon Raman laser. *Nature* **433**, 292–294 (2005).
- Liang, T. K. & Tsang, H. K. Role of free carriers from two-photon absorption in Raman amplification in silicon-on-insulator waveguides. *Appl. Phys. Lett.* **84**, 2745–2747 (2004).
- Rong, H. *et al.* Raman gain and nonlinear optical absorption measurement in a low loss silicon waveguide. *Appl. Phys. Lett.* **85**, 2196–2198 (2004).
- Claps, R., Raghunathan, V., Dimitropoulos, D. & Jalali, B. Role of nonlinear absorption on Raman amplification in Silicon waveguides. *Opt. Expr.* **12**, 2774–2780 (2004).
- Agrawal, G. P. *Nonlinear Fiber Optics* 2nd edn, 42–43 (Academic, New York, 1995).
- Jones, R. *et al.* Net continuous wave optical gain in a low loss silicon-on-insulator waveguide by stimulated Raman scattering. *Opt. Expr.* **13**, 519–525 (2005).
- Siegman, A. E. *Lasers* 413–418 (University Science Books, Mill Valley, 1986).
- Suto, K., Kimura, T. & Nishizawa, J. Semiconductor Raman laser pumped with a fundamental mode. *IEEE Proc. J.* **139**, 407–412 (1992).
- Brasseur, J. K., Repasky, K. S. & Carlsten, J. L. Continuous-wave Raman laser in H_2 . *Opt. Lett.* **23**, 367–369 (1998).
- Garmire, E. Criteria for optical bistability in a lossy saturating Fabry-Pérot. *IEEE J. Quant. Electron.* **QE-25**, 289–295 (1989).
- Salib, M. *et al.* Silicon photonics. *Intel Technol. J.* **8**, 143–160 (2004).
- Liu, M. Y. & Chou, S. Y. High-modulation-depth and short-cavity-length silicon Fabry-Pérot modulator with two grating Bragg reflectors. *Appl. Phys. Lett.* **68**, 170–172 (1996).
- Barrios, C. A., Almeida, V. R., Panepucci, R. R., Schmidt, B. S. & Lipson, M. Compact silicon tunable Fabry-Pérot resonator with low power consumption. *IEEE Photon. Technol. Lett.* **16**, 506–508 (2004).
- Headley, W. R., Reed, G. T., Liu, A., Paniccia, M. & Howe, S. Polarization independent optical racetrack resonators using rib waveguide on silicon on insulator. *Appl. Phys. Lett.* **85**, 5523–5525 (2004).
- Lee, M.-C. M. & Wu, M. C. in *IEEE/LEOS Int. Conf. on Optical MEMS and Their Applications (Hawaii, 18–21 August 2003)* Paper MC3 28–29 (IEEE, Piscataway, NJ, 2003).
- Koonath, P., Indukuri, T. & Jalali, B. Vertically coupled micro-resonators realized using three-dimensional sculpting in silicon. *Appl. Phys. Lett.* **85**, 1018–1020 (2004).
- Kippenberg, T. J., Spillane, S. M., Armani, D. K. & Vahala, K. J. Ultralow-threshold microcavity Raman laser on a microelectronic chip. *Opt. Lett.* **29**, 1224–1227 (2004).

Acknowledgements We thank R. Nicolaescu for initiating and developing this research through the early stages; A. Alduino, D. Tran, J. Tseng, A. Barkai and D. Hodge, for assistance in device fabrication and sample preparation; S. Koehl for software development; L. Premislov for scanning electron microscope images; M. Morse, H. Liu, M. Salib, D. Samararubio, L. Liao, R. Li and G. Ding, for technical discussions; G. T. Reed and I. P. Kaminow for helpful conversations.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.R. (haisheng.rong@intel.com).

Biological control of terrestrial silica cycling and export fluxes to watersheds

Louis A. Derry¹, Andrew C. Kurtz², Karen Ziegler³ & Oliver A. Chadwick³

¹Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York 14853, USA

²Department of Earth Sciences, Boston University, Boston, Massachusetts 02215, USA

³Department of Geography, University of California, Santa Barbara, California 93106, USA

Silicon has a crucial role in many biogeochemical processes—for example, as a nutrient for marine and terrestrial biota, in buffering soil acidification and in the regulation of atmospheric carbon dioxide. Traditionally, silica fluxes to soil solutions and stream waters are thought to be controlled by the weathering and subsequent dissolution of silicate minerals^{1,2}. Rates of mineral dissolution can be enhanced by biological processes³. But plants also take up considerable quantities of silica from soil solution, which is recycled into the soil from falling litter in a separate soil–plant silica cycle that can be significant in comparison with weathering input and hydrologic output^{4–8}. Here we analyse soil water in basaltic soils across the Hawaiian islands to assess the relative contributions of weathering and biogenic silica cycling by using the distinct signatures of the two processes in germanium/silicon ratios. Our data imply that most of the silica released to Hawaiian stream water has passed through the biogenic silica pool, whereas direct mineral–water reactions account for a smaller fraction of the stream silica flux. We expect that other systems exhibiting strong Si depletion of the mineral soils and/or high Si uptake rates by biomass will also have strong biological control on silica cycling and export.

Germanium/silicon ratios have been used to trace silica sources in rivers and the oceans^{9,10}. Ge/Si ratios in streams unaffected by pollution or hydrothermal inputs are always lower than Ge/Si in the silicate bedrock that they drain¹¹. Secondary minerals (clays) formed in the soil environment are the major complementary reservoir higher in Ge/Si (refs 12–14). Dissolved Ge–Si relationships from most rivers from a variety of climatic and geological settings are similar, and comprise trends that have been explained as mixing between silica derived from weathering of primary minerals and silica derived from weathering of secondary clays^{12,13}. According to this scheme, which we term the Murnane, Stallard, Froelich (MSF) model, incongruent dissolution of primary minerals yields a solution with high Si concentration, [Si], and low Ge/Si (component 1), whereas dissolution of secondary, Ge-enriched clays yields a low-[Si], high-Ge/Si solution (component 2).

The MSF model assumes direct mineralogical control on stream silica (and germanium). According to this model, mineral stability

controls the concentrations of dissolved silica in equilibrium with clay assemblages; mineral weathering rates define the available flux of Si; and the degree of dilution of solutions generated in the weathering environment sets overall stream [Si]. Ge/Si ratios in soil solutions should reflect the soil mineral phases undergoing dissolution or transformation¹⁵. However, the recognition of an active terrestrial biological cycle of Si raises questions about the pathways of silica from weathering to streams. Many plants sequester silica in biogenic phytoliths (opal-A structures), and soils can accumulate significant quantities of biogenic opal-A¹⁶. What is the impact of this internal plant cycle on stream Si fluxes? Is stream export largely unaffected by this process, or does the 'internal' plant cycle have a significant role in controlling silica export from watersheds?

The Hawaiian islands offer an opportunity to test the predictions of the MSF model and to investigate the impact of biogenic silica cycling on stream export. Ge/Si in fresh Hawaiian basalts (tholeiitic and alkaline) fall in a relatively narrow range of $(2.3\text{--}2.9) \times 10^{-6}$ (mol/mol) (ref. 17). Ge and Si data from Hawaiian streams unaffected by coal burning or hydrothermal inputs define an apparent mixing relationship¹¹ (Fig. 1). The data can be modelled as a two-component mixture in which component 1 has $\text{Ge/Si} \approx 0.2 \times 10^{-6}$ and $[\text{Si}] > 600 \mu\text{M}$, and component 2 has $\text{Ge/Si} \approx 2.6 \times 10^{-6}$ and $[\text{Si}] \leq 25 \mu\text{M}$ (Fig. 1). The MSF model predicts that the low-Ge/Si, high-[Si] component 1 is produced during the incongruent dissolution of primary minerals in the lower part of soil profiles, near the interface with fresh material, and/or in the upper parts of young soils that are not yet depleted of silica and retain primary minerals. High-Ge/Si, low-[Si] solutions (component 2) should occur in the upper parts of weathered soil profiles that have lost their primary minerals and in which the dissolution of secondary clays with high Ge/Si is important.

We tested the predictions of the above model by measuring [Ge] and [Si] in soil waters from seven depth profiles along a chrono-sequence of basaltic soils across the Hawaiian islands. The chrono-sequence ranges from 0.3 to 4,100 kyr substrate age, but has common rock type, present-day climate and vegetation^{18,19}. Major element geochemistry, soil mineralogy and solid-phase Ge/Si ratios have been characterized by previous studies. Young soils (about 0.3 kyr old) retain primary minerals and volcanic glass, have experienced little Si loss, and have Ge/Si ratios close to that of fresh basalt¹⁴. Older soils (at least 20 kyr old) have experienced

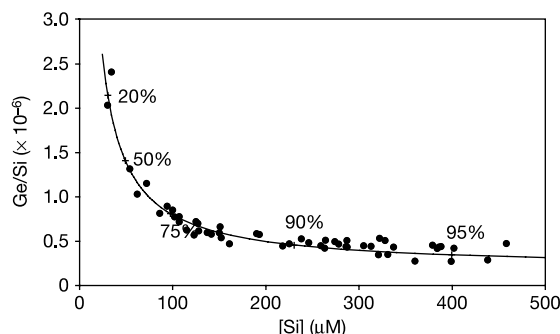


Figure 1 Plot of Ge/Si against [Si] from Hawaiian stream waters. Data are from ref. 12 and this study. The hyperbolic curve is a two-component mixing model, in which component 1 is derived from the dissolution of biogenic silica and has $\text{Ge/Si} = 0.25 \times 10^{-6}$ and $[\text{Si}] = 1,800 \mu\text{M}$, and component 2 is derived from the dissolution of soil minerals and has $\text{Ge/Si} = 2.6 \times 10^{-6}$ and $[\text{Si}] = 25 \mu\text{M}$. Hawaiian streams generally show lower [Si] and higher Ge/Si at high discharge. The discharge-weighted stream fluxes have $\text{Ge/Si} < 1$ and $[\text{Si}] > 80 \mu\text{M}$, implying that component 1 is the larger contributor of dissolved silica and indicating a major role for biogenic silica in controlling stream Si export. Labelled crosses on the mixing curve indicate the modelled fraction of biogenic Si.

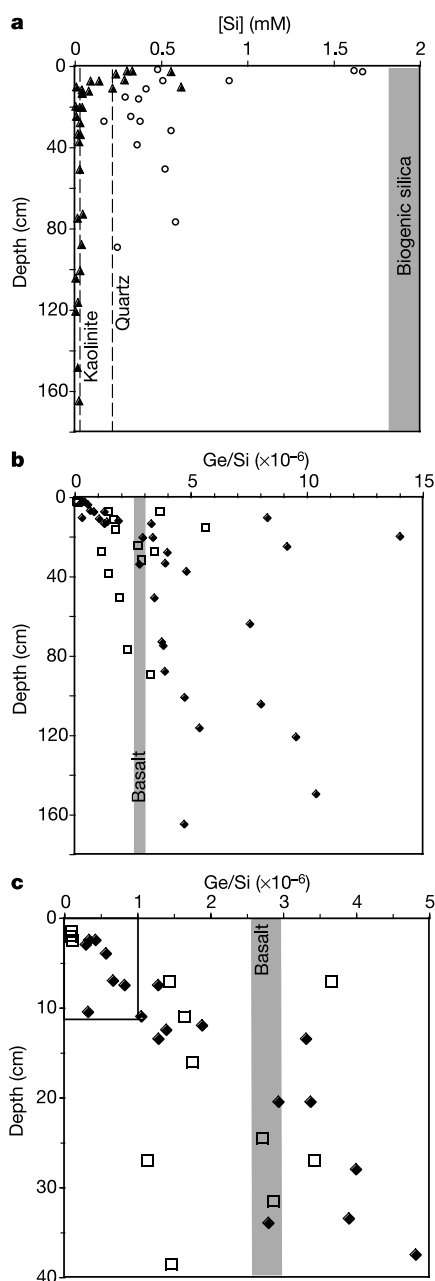


Figure 2 Dissolved silica and Ge/Si ratios from Hawaiian soil waters. **a**, Dissolved silica from Hawaiian soil profiles. Solutions from older (at least 20 kyr old; filled triangles), highly weathered soils have low $[Si]_{sol}$, near equilibrium with gibbsite–kaolinite. Higher values are found only in the upper 15 cm of the soil profile. Solutions from young soils (less than 5 kyr old; open circles) have higher dissolved $[Si]_{sol}$ at all depths, indicative of weathering of primary minerals and glass. The highest $[Si]_{sol}$ values are from the upper 15 cm and in some cases approach opal saturation. Vertical dashed lines are for reference and indicate saturation with kaolinite and quartz. The grey field represents an estimate of the solubility of biogenic opal^{21,29}, although some temperate species may show lower solubilities³⁰. **b**, Dissolved Ge/Si ratios from older soils (filled diamonds) are high (more than 2×10^{-6}) except above 15 cm. $(Ge/Si)_{sol}$ values from young soils (open squares) are generally lower than those in the old soils but are more than 1×10^{-6} except above 10 cm. $(Ge/Si)_{sol}$ values are mostly low (less than 0.5×10^{-6}) in both young and old soils in the upper 10 cm. **c**, An enlarged view of a portion of **b**. The rectangle delimits the region in which samples with $Ge/Si < 1 \times 10^{-6}$ occur. Samples with both high dissolved $[Si]$ and low Ge/Si necessary to produce the observed stream values are found only in the upper 10–15 cm of soils, regardless of soil age.

significant Si loss, and have high Ge/Si (about 6×10^{-6} to 20×10^{-6}). Allophane is the dominant secondary mineral in soils from 20 to 150 kyr, whereas the oldest sites (1,400 and 4,100 kyr) contain kaolin minerals, gibbsite and sesquioxides¹⁸.

Soil solutions were extracted by the saturation paste method²⁰ and analysed for Si and Ge (details are given in Methods). At sites at least 20 kyr old, soil solutions from deeper than 15 cm range from 6 to $45 \mu M$ Si, near the value expected for kaolinite–gibbsite equilibrium (about $30 \mu M$). Soil solution, $[Si]_{sol}$, values more typical of concentrations in streams ($100 \mu M$ or more) are only present in the upper 15 cm (Fig. 2a). In young (about 0.3 kyr old) sites, $[Si]_{sol}$ is mostly 200–600 μM , a value consistent with active weathering of fresh glass and primary minerals throughout the soil profiles. However, in near-surface samples, $[Si]_{sol}$ is very high and can approach experimentally determined values of phylolith solubility²¹. Ge/Si ratios in the soil waters are lowest in near-surface organic horizons or those immediately below, and are much higher in underlying mineral soils (Fig. 2b). $(Ge/Si)_{sol} < 1 \times 10^{-6}$ are found only in the upper 10 cm at all sites (and only in the upper 5 cm at the young sites), and are associated with relatively high $[Si]_{sol}$. Below 15 cm, $(Ge/Si)_{sol}$ in the old soils almost always exceeds the fresh basalt value ($2.6 \mu mol mol^{-1}$), ranging from near 3×10^{-6} to 14×10^{-6} . The highest values are found at the oldest site, and are associated with very low $[Si]_{sol}$ values: 10 μM or less. In the young soils, $(Ge/Si)_{sol}$ below 10 cm range from 1.1×10^{-6} to 5.6×10^{-6} , bracketing the value of fresh basalt.

High- $(Ge/Si)_{sol}$, low- $[Si]_{sol}$ soil waters found below 15 cm depth at the older sites are consistent with the prediction of the MSF model: a ‘component 2’ composition resulting from the dissolution of Ge-enriched secondary aluminosilicates. At the young sites (less than 20 kyr old), where primary minerals and glass are still present and undergoing weathering, $(Ge/Si)_{sol}$ is mostly less than or equal to $(Ge/Si)_{basalt}$, which is consistent with the production of a Ge-depleted solution during primary mineral weathering. However, except in the top 5-cm organic horizon, $(Ge/Si)_{sol}$ values are too high to account for the low- Ge/Si component 1 required by stream data. Primary basaltic weathering does not provide this low- Ge/Si component. In contrast to the predictions of the MSF model, a ‘component 1’ composition with $Ge/Si < 0.5 \times 10^{-6}$ and high $[Si]$ is found only in near-surface (organic) horizons in both young and old soil profiles.

High- $[Si]$, low- (Ge/Si) values in the surface soil water samples are consistent with a biogenic input of Si (refs 5, 6). We extracted phyloliths from leaf samples across the soil chronosequence (see details in Methods section). *Cybotium*, *Dicraopteris* and *Diplazium* ferns contain significant amounts of phylolith Si, whereas the dominant canopy species *Metrosideros polymorpha* does not. Ge/Si values in the phyloliths were low, from 0.04×10^{-6} to 0.37×10^{-6} , similar to the low $(Ge/Si)_{sol}$ values from the surface soils and to component 1 of the stream model (Fig. 1). This biogenic component is the only one we have identified that can provide the low- Ge/Si , high- $[Si]$ component required to explain the data from Hawaiian streams.

Given the constraints described above, we can apply a mixing model to determine the relative importance of the two Si sources to four Hawaiian streams gauged by the US Geological Survey, which were sampled by Mortlock and Froelich¹¹. Combined stream, soil water and phylolith data suggest a value of 0.2 ± 0.1 for component 1. $[Si]$ for this component must be at least 1,600 μM (observed in soil solutions from the Ola’a site) but cannot exceed the solubility of phyloliths (about 1,800 μM (ref. 21)). A best fit to a linearized projection of the stream data in Fig. 1 yields an estimate ($\pm s.e.$) of $Ge/Si = 0.25 \pm 0.02$ at $[Si] = 1,800 \mu M$ (see Supplementary Information), and we take these values as representative of component 1 (phylolith-derived Si) for modelling purposes. Combined stream and soil water data suggest $[Si]_{sol} = 25 \mu M$ for component 2 (weathering-derived Si), with a Ge/Si ratio of 2.6×10^{-6} . The

discharge-weighted mean Ge/Si values from the four streams range from 0.49 to 0.93 and indicate that component 1 comprises 68–90% of Si carried by stream water. Given the high [Si] of this component, a concentration-weighted calculation requires that only 2.8–9% of the stream water flux is derived from this shallow source. We note that a lower [Si]_{sol} value for either of components 1 or 2 (with higher [Ge]) could still provide a good fit to the data (Fig. 1) but would increase the estimated biogenic silica fraction. Conversely, inadequate sampling of high discharge events with low [Si] and high Ge/Si could bias the mean values towards a biogenic source.

Silica supply to Hawaiian soils is controlled by weathering of the local basaltic substrate and mineral aerosols, which contribute significant amounts of Si to older soils²². The export of silica from the watersheds cannot routinely exceed the supply. Export fluxes from the four watersheds with long-term data used here are 150–5,400 mol Si ha⁻¹ yr⁻¹. Silica export fluxes from the rooting zone (20–30 cm depth) have been measured for our soil sites, and range from 400–9,400 mol Si ha⁻¹ yr⁻¹ (ref. 23), so the production of dissolved Si from shallow soils seems capable of supporting our inferred biogenic fraction of watershed Si losses. Although data on the rates of uptake of Si from soils by vegetation are not yet available, it is very likely that the internal cycle of Si is much faster than the rate of external input. We propose that silica is extensively recycled between the upper soil profile and plants, as are other nutrients, and dissolved silica in the shallow soils acquires a low Ge/Si ratio typical of phytolith silica.

Our data imply that most of the silica released to Hawaiian stream water has passed through the biogenic silica pool, whereas direct mineral–water reactions account for a smaller fraction of the stream Si flux. Although mineral dissolution reactions must ultimately control the availability of silica in the environment, in Hawaii the biological cycle of silica seems to be the dominant control on the rate of silica loss from watersheds. We suggest that amorphous biogenic silica in the near-surface soil is much more labile than Si present in clays in underlying mineral horizons, or even in primary mineral assemblages²¹.

The data we have presented here argue for a strong biological imprint on the silica cycle in a mesic tropical, volcanic system. Hawaiian soils are generally silica-depleted, leaving the biogenic Si pool with an important role in the overall silica cycle. The plant cycle of Si maintains high dissolved [Si] in soil waters in the upper portion of soils at highly weathered sites that would otherwise have [Si] values at least an order of magnitude lower. A substantial fraction of root activity and nutrient uptake takes place in this shallow rooting zone at the Hawaiian sites²⁴, and high concentrations of dissolved silica might act as a buffer to otherwise high and potentially toxic dissolved aluminium concentrations in the rooting zone²⁵. Hawaii should not be unique in this way, because other highly weathered tropical systems share the characteristics of strong Si depletion of the mineral soils and high biomass production and turnover, creating a potentially large pool of reactive biogenic silica relative to low availability of mineral-derived Si. Under these conditions the biogenic Si pool will be important in the silica cycle^{4,26}. We expect that some other terrestrial ecosystems, including grasslands that are rich in biogenic silica and have high turnover rates^{8,27}, also have a strong biological control on Si cycling and export. By controlling the activity of dissolved Si in soils and streams, plants might exert widespread influence on weathering rates and terrestrial silica fluxes. □

Methods

Water samples

Stream water samples were filtered through 0.22-µm filters, spiked with a ⁷⁰Ge tracer solution, and analysed by hydride-generation inductively coupled plasma (ICP) mass spectrometry on a Finnigan Element 2 at Cornell University²⁸. Soil samples were taken as continuous channel samples from hand-dug soil pits; individual samples therefore represent a depth range of 4–30 cm. Soil water samples were prepared by mixing distilled

water with homogenized soil samples to field capacity. This standard method provides an appropriate mimic for field conditions because most soil water is held at or below field capacity and is then flushed by occasional rainfall events. With the exception of short periods of saturated flow, nearly all water in soil is subject to unsaturated flow for long periods before leaching. In time-series extraction studies we found that [Si]_{sol} reached constant values within several hours. After 48 h of equilibration the samples were centrifuged at 5,000 r.p.m. The supernatant was filtered at 0.22 µm, then spiked and analysed as above. Silica concentrations were determined by ICP optical emission spectroscopy. Uncertainties in both Ge and Si measurements were less than 5%.

Phytolith extraction

Opal phytoliths were extracted from leaf samples by closed-vessel microwave digestion in a Milestone Ethos Plus system. Dry leaf sample (0.5 g) was placed in a 100-ml Teflon digestion vessel with 9 ml of HNO₃ and 1 ml of H₂O₂. Care was taken to make sure that the sample was wetted, and the sample was left uncovered for 10 min. It was then capped, placed in the microwave, and heated to 180 °C for 10 min. After digestion, the sample was centrifuged and the acid was decanted from the opal. After several rinses with deionized water, the opal was digested at room temperature in 2 M NaOH. The digested sample was diluted and analysed for Ge and Si at Boston University by procedures similar to those described above.

Received 9 June; accepted 15 December 2004; doi:10.1038/nature03299.

- Garrels, R. M. in *Researches in Geochemistry* (ed. Alabon, P. H.) 405–420 (Wiley, New York, 1967).
- Drever, J. I. *The Geochemistry of Natural Waters* (Prentice-Hall, Englewood Cliffs, NJ, USA, 1988).
- Moulton, K. L., West, J. & Berner, R. A. Solute flux and mineral mass balance approaches to the quantification of plant effects on silicate weathering. *Am. J. Sci.* **300**, 539–570 (2000).
- Lucas, Y., Luizao, F. J., Chauvel, A., Rouiller, J. & Nahon, D. The relation between biological activity of the rain forest and mineral composition. *Science* **260**, 521–523 (1993).
- Alexandre, A., Meunier, J.-D., Colin, F. & Koud, J.-M. Plant impact on the biogeochemical cycle of silicon and related weathering processes. *Geochim. Cosmochim. Acta* **61**, 677–682 (1997).
- Meunier, J. D., Colin, F. & Alarcon, C. Biogenic silica storage in soils. *Geology* **27**, 835–838 (1999).
- Markewitz, D. & Richter, D. D. The bio in aluminum and silicon biogeochemistry. *Biogeochemistry* **42**, 235–252 (1998).
- Conley, D. J. Terrestrial ecosystems and the global biogeochemical silica cycle. *Global Biogeochem. Cycles* **16**, doi:10.1029/2002GB001894 (2002).
- Froelich, P. N. & Andreae, M. O. The marine geochemistry of germanium: ekasilicon. *Science* **213**, 205–207 (1981).
- Froelich, P. N., Hambrick, G. A., Andreae, M. O., Mortlock, R. A. & Edmond, J. M. The geochemistry of germanium in natural waters. *J. Geophys. Res.* **90**, 1133–1141 (1985).
- Mortlock, R. A. & Froelich, P. N. Continental weathering of germanium: Ge/Si in the global river discharge. *Geochim. Cosmochim. Acta* **51**, 2075–2082 (1987).
- Murnane, R. J. & Stallard, R. F. Germanium and silicon in rivers of the Orinoco drainage basin. *Nature* **344**, 749–752 (1990).
- Froelich, P. N. *et al.* River fluxes of dissolved silica to the ocean were higher during the glacials: Ge/Si in diatoms, rivers, and oceans. *Paleoceanography* **7**, 739–768 (1992).
- Kurtz, A. C., Derry, L. A. & Chadwick, O. A. Germanium/silicon fractionation in the weathering environment. *Geochim. Cosmochim. Acta* **66**, 1525–1537 (2002).
- Kurtz, A. C. & Derry, L. A. in *Proc. 11th Int. Symp. Water Rock Interaction* (eds Wanty, R. B. & Seal, R. R.) 833–837 (Swets & Zeitlinger, Lisse, The Netherlands, 2004).
- Drees, L. R., Wilding, L. P., Smek, N. E. & Senkay, A. L. in *Minerals in Soil Environments* 2nd edn. (eds Dixon, J. B. & Weed, S. B.) 914–974 (Soil Science Soc. Am., Madison, Wisconsin, 1989).
- DeArgollo, R. & Schilling, J.-G. Ge–Si and Ga–Al fractionation in Hawaiian volcanic rocks. *Geochim. Cosmochim. Acta* **42**, 623–630 (1978).
- Vitousek, P. M. *et al.* Soil and ecosystem development across the Hawaiian Islands. *GSA Today* **7**, 1–10 (1997).
- Chadwick, O. A., Derry, L. A., Vitousek, P. M., Huebert, B. J. & Hedin, L. O. Changing sources of nutrients during four million years of ecosystem development. *Nature* **397**, 491–497 (1999).
- Lajtha, K., Jarrell, W. M., Johnson, D. W. & Sollins, P. in *Standard Soil Methods for Long-Term Ecological Research* (eds Robertson, G. P., Coleman, D. C., Bledsoe, C. S. & Sollins, P.) 166–182 (Oxford Univ. Press, New York, 1999).
- Frayse, F., Pokrovsky, O. S., Schott, J. & Meunier, J. D. Surface properties, solubility and dissolution kinetics of phytoliths from bamboos of Reunion Island. *Geochim. Cosmochim. Acta Suppl. (Proc. 14th Ann. Goldschmidt Conf.)*, A216 (2004).
- Kurtz, A. C., Derry, L. A. & Chadwick, O. A. Accretion of Asian dust to Hawaiian soils: isotopic, elemental and mineral mass balances. *Geochim. Cosmochim. Acta* **65**, 1971–1983 (2001).
- Hedin, L. O., Vitousek, P. & Matson, P. Nutrient losses over four million years of tropical forest development. *Ecology* **84**, 2231–2255 (2003).
- Vitousek, P. M., Gerrish, G., Turner, D. R., Walker, L. R. & Mueller-Dombois, D. Litterfall and nutrient cycling in 4 Hawaiian montane rain-forests. *J. Trop. Ecol.* **11**, 189–203 (1995).
- Exley, C., Schneider, C. & Doucet, F. J. The reaction of aluminium with silicic acid in acidic solution: an important mechanism in controlling the biological availability of aluminium? *Coord. Chem. Rev.* **228**, 127–135 (2002).
- Meunier, J. D. The role of plants in the terrestrial biogeochemical cycle of Si. *Geochim. Cosmochim. Acta Suppl. (Proc. 14th Ann. Goldschmidt Conf.)*, A411 (2004).
- Kelly, E. F., Chadwick, O. A. & Hilinski, T. E. The effect of plants on mineral weathering. *Biogeochemistry* **42**, 21–53 (1998).
- Mortlock, R. A. & Froelich, P. N. Determination of germanium by isotope dilution–hydride generation inductively coupled plasma mass spectrometry. *Anal. Chem.* **332**, 277–284 (1996).
- Dove, P. M. in *Chemical Weathering Rates in Silicate Minerals* (eds White, A. F. & Brantley, S. L.) 235–290 (Mineralogical Society of America, Washington DC, 1995).
- Bartoli, F. & Wilding, L. P. Dissolution of biogenic opal as a function of its physical and chemical properties. *Soil Sci. Soc. Am. J.* **44**, 873–878 (1980).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Moore and P. T. Atkins for field and laboratory assistance. This research was supported by grants from the A. W. Mellon Foundation to L.A.D. and O.A.C., and from the N.S.F. to L.A.D., O.A.C. and A.C.K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.A.D. (lad9@cornell.edu).

Pressure sensitivity of olivine slip systems and seismic anisotropy of Earth's upper mantle

David Mainprice¹, Andr  a Tommasi¹, H  l  ne Couvy^{2,3}, Patrick Cordier² & Daniel J. Frost³

¹Laboratoire de Tectonophysique, CNRS/Universit   de Montpellier II, F-34095 Montpellier cedex 5, France

²Laboratoire Structure et Propri  t  s de l'Etat Solide, CNRS/Universit   de Lille I, F-59650 Villeneuve d'Ascq, France

³Bayerisches Geoinstitut, Universit  t Bayreuth, D-95440 Bayreuth, Germany

The mineral olivine dominates the composition of the Earth's upper mantle and hence controls its mechanical behaviour and seismic anisotropy. Experiments at high temperature and moderate pressure, and extensive data on naturally deformed mantle rocks, have led to the conclusion that olivine at upper-mantle conditions deforms essentially by dislocation creep with dominant [100] slip. The resulting crystal preferred orientation has been used extensively to explain the strong seismic anisotropy observed down to 250 km depth^{1–4}. The rapid decrease of anisotropy below this depth has been interpreted as marking the transition from dislocation to diffusion creep in the upper mantle⁵. But new high-pressure experiments suggest that dislocation creep also dominates in the lower part of the upper mantle, but with a different slip direction. Here we show that this high-pressure dislocation creep produces crystal preferred orientations resulting in extremely low seismic anisotropy, consistent with seismological observations below 250 km depth. These results raise new questions about the mechanical state of the lower part of the upper mantle and its coupling with layers both above and below.

Despite the considerable effort to characterize olivine's deformation mechanisms over the past 30 yr, it is only recently that deformation experiments could be conducted at pressure–temperature conditions of the entire upper mantle^{6–8}. New simple-shear experiments on olivine aggregates at 11 GPa and 1,400   C, conditions equivalent to those at depths of 330 km, have shown that deformation takes place by dislocation creep, with dominant activation of [001]{hk0} slip systems⁹, suggested by the concentration of [001] parallel to the shear direction and of [100] and [010] normal to the shear plane (Fig. 1). Transmission electron microscopy shows the exclusive presence of dislocations with [001] Burgers vectors in a screw orientation, compatible with [001](hk0) slip. Dominant [001] slip in the deep upper mantle requires re-evaluation of the interpretation of anisotropic physical properties. For instance, the fastest P-wave velocity will no longer parallel the shear direction as in an upper mantle deforming by dominant [100](010) slip, which is the assumption traditionally used in relating flow and seismic anisotropy in the mantle^{10,11}.

Several lines of evidence point to seismic anisotropy decreasing with depth in the upper mantle. Most global one-dimensional models (PREM, IASP, AK135 and AK303) show horizontally propagating P waves travelling (at velocity v_{PH}) faster than vertical ones (at v_{PV}), but the difference in velocity reduces with depth, resulting in isotropic behaviour at 350 km depth⁴. Some models (AK135 and 303) even show v_{PV} slightly faster than v_{PH} below 350 km. The S-wave polarization anisotropy also decreases monotonically from the surface to become isotropic at 250 km. For horizontally propagating S waves, horizontally polarized waves show a higher velocity (v_{SH}) than vertically polarized waves (v_{SV}) down to about 250 km depth. Between 300 km and 400 km depth, v_{SV} is higher than v_{SH} , but anisotropy is five times lower than in the uppermost mantle. High-resolution global tomographic models based on S-wave data¹² or on the inversion of three-component surface and body waveform data¹³ support these general findings, with strong anisotropy characterized by $v_{SH} > v_{SV}$ above 250 km depth. At greater depth, these models require a strong decrease in anisotropy, with a minimum around 300 km depth. S-wave data also call for weak anisotropy, with $v_{SV} > v_{SH}$ at the base of the upper mantle beneath the central Pacific and Pre-Cambrian cratons¹². Regional surface wave studies in the Pacific and Indian ocean basins also suggest that anisotropy is present from the surface to ~250–300 km depth^{1–3,14}, with v_{SH} being greater than v_{SV} . Analysis of two-station surface wave profiles in the Pacific and Philippine plates imply a still shallower anisotropy limited to the upper 160 km of the mantle¹⁵. SKS studies cannot constrain the depth of the anisotropic layer, but the strong correlation of the direction of polarization of the fast shear wave with the surface geology and the observed delay times ≤ 2 s (ref. 14) suggests that SKS splitting occurs in the upper 200–250 km of the mantle.

Finally, a regional seismic discontinuity, called the Lehmann discontinuity, has been detected at about 220 km depth by various seismic methods (reflection, surface waves, ScS reverberations and P to S conversions), mainly beneath continents. This discontinuity has been interpreted as being due to either (1) a strong anisotropy caused by intense deformation of olivine in a zone of mechanical coupling between the lithosphere and the asthenosphere¹⁶, or (2) the transition between an anisotropic uppermost mantle deforming by dislocation creep (which produces a crystal preferred orientation, CPO, of olivine) and an isotropic deep mantle deforming by diffusion creep (which does not produce CPO)⁵. However, recent high-pressure, high-temperature experiments show that even in fine-grained aggregates (~20–30   m), dislocation creep is the dominant deformation mechanism under conditions equivalent to those prevailing at 300 km depth^{9,17,18}.

The pressure, or pressure interval, at which the transition from

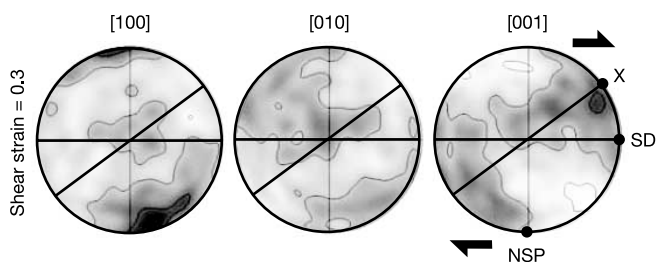


Figure 1 Preferred orientation of [100], [010] and [001] crystallographic axes in synthetic olivine polycrystal S2954 deformed at 1,400   C and 11 GPa confining pressure in simple shear⁹. Lower hemisphere equal-area projection, contours at intervals of 0.5 multiples of a uniform distribution. 3,269 measured orientations. Dextral shear (top to the right) is indicated by half-arrows; SD, shear direction; NSP, normal to shear plane; X, finite strain extension direction. Shear strain ~0.3. Inclined black line marks the foliation (flattening plane).

[100](010) to [001]($hk0$) slip occurs has still not been characterized, but it must be less than 11 GPa, which corresponds to 330 km depth⁹. Constraints on the minimum depth for this transition may be derived from the analysis of olivine CPO patterns in naturally deformed mantle rocks¹⁹. Spinel-peridotites that are equilibrated above 70 km depth display solely CPO characteristic of [100] slip. This CPO is also dominant in high-pressure garnet-peridotites from South Africa equilibrated between 70 km and 150 km depth²⁰. CPO suggesting activation of both slip directions at high-temperature conditions is restricted to rare high-pressure peridotite mylonites from the Tanzanian and Kaapvaal cratons equilibrated at ~140 km depth²¹. This suggests that [100] slip dominates in the mantle above 150 km depth.

We simulated development of CPOs in olivine polycrystals deformed in simple shear under high-pressure conditions, using a viscoplastic self-consistent (VPSC) model²² that has been extensively tested for olivine^{23,24}. In this model, as in all polycrystal plasticity approaches, CPO evolution is essentially controlled by the imposed deformation, the initial texture, and the active slip systems. The last depend on the mineral structure, but also on the temperature and pressure conditions, which control their relative strength or critical resolved shear stress (CRSS). In Fig. 2, we show the CPO developed in an aggregate of 500 initially spherical and randomly oriented olivine grains after a shear strain of 0.3 and 1.0; [001] axes tend to align in-between the stretching and the shear direction, and [100] and [010] axes concentrate at high angles to the shear plane. In this simulation, CRSS for the [001]($hk0$), [100](001) and [100](010), and [100](011) and [100](021) systems are 1:3:6; that is, slip on [001]($hk0$) is three times easier than slip on [100](010) and six times easier than on [100](021). Yet tests with different CRSS values predict similar CPO for all combinations in which slip in [001] systems is significantly easier than in [100] systems. Comparison with olivine CPO formed in recent high-temperature high-pressure experiments in simple shear (Fig. 1) provides evidence that these models give good estimates of olivine CPO in the lower part of the upper mantle.

The three-dimensional distribution of seismic velocities in a polycrystalline aggregate may be estimated by averaging the individual grain elastic constant tensors as a function of the crystallographic orientations and mineralogical composition of the aggregate. Seismic properties (Fig. 3) of an upper-mantle sample with pyrolitic composition (63% olivine, 17% garnet, 20% clino-

pyroxene) at a pressure of 11.8 GPa and temperature of 1,380 °C, corresponding to 355 km depth, were calculated using recently determined elastic constant tensors of olivine²⁵, pyrope-rich garnet²⁶ and diopside²⁷. Olivine displays the modelled CPO (Fig. 2). Garnet has a random CPO; this agrees with predictions of VPSC simulations and observations in naturally deformed garnet-rich rocks that show that garnet CPOs are always very weak²⁸. We have assumed that diopside also has a random orientation as high-pressure data are lacking for this mineral. However, if diopside has a CPO similar to that developed under high-temperature/low-pressure conditions²⁹, it will tend to reduce anisotropy by destructive interference with olivine³⁰. Both compressional and shear waves display weak anisotropies (0.9% and 1.9%, respectively). The fastest compressional waves propagate at a high angle to the shear plane. The variation of compressional waves' velocities within the shear plane is very small. Thus in a mantle deforming by horizontal shearing, almost no azimuthal variation of P-wave velocity (v_P) would be observed and vertically propagating P waves would be only slightly faster than those propagating horizontally. The polarization anisotropy of shear waves is characterized by faster propagation of waves polarized at a high angle to the shear plane, and the largest delay times are observed for propagation at a high angle to the shear direction in the shear plane. For propagation in the shear plane, the fastest S waves are polarized at a high angle to the shear plane. Hence, for horizontal flow, v_{SV} is greater than v_{SH} .

All the above predictions are consistent with global and regional seismic observations^{1-4,12-14}, which show a weak anisotropy (<2%) below 250 km depth for P and S waves. The predicted S-wave anisotropy, although weak (1.9%), is twice as strong as the P-wave anisotropy observed in global models⁴. Anisotropy patterns observed in global models (transverse isotropy with a vertical symmetry axis) are best reproduced for horizontal shearing. For such a flow pattern, our simulations predict that vertical P waves propagate faster than horizontal ones, and that for horizontally propagating S waves v_{SV} is greater than v_{SH} , in agreement with global models of anisotropy patterns for depths greater than 300 km. In addition, vertically propagating shear waves will detect no anisotropy.

We conclude that dominant activity of [001]($hk0$) in olivine at high pressure is entirely compatible with the available seismic data, which indicate a weak anisotropy in the upper mantle below 300 km. It is difficult to imagine another scenario that would reproduce the anisotropy patterns of P and S waves in such detail. The interpretation presented here allows us to consider that the

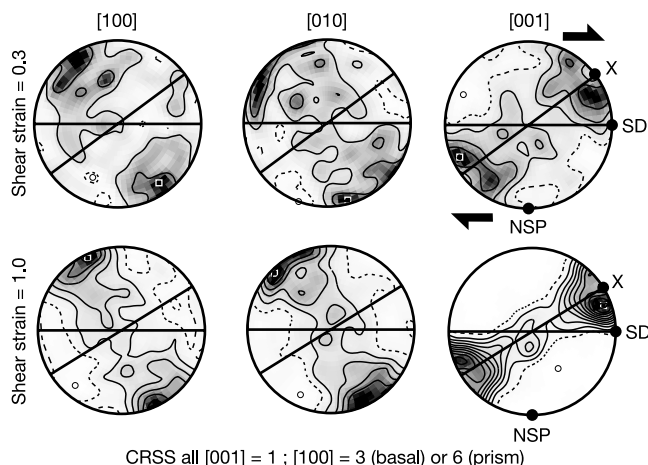


Figure 2 Olivine crystal preferred orientations predicted using a viscoplastic self-consistent model. Data are shown for a shear strain of 0.3 (top) and 1.0 (bottom). Lower hemisphere equal-area projection, contours at intervals of 0.5 multiples of a uniform distribution, 500 grains. Symbols and abbreviations as Fig. 1.

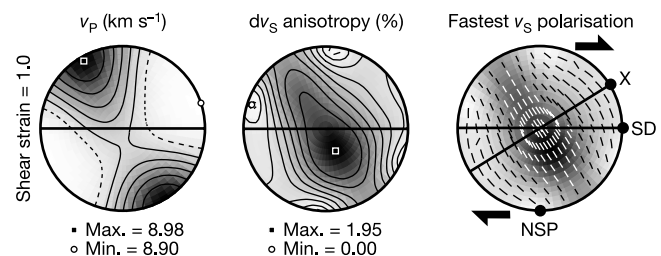


Figure 3 Modelled three-dimensional compressional velocity and shear wave anisotropy distributions, and fastest shear wave polarization. Data shown for an aggregate composed of 63% anisotropic olivine and isotropic garnet and diopside (17% and 20%, respectively) at 11.8 GPa and 1,380 °C. Lower hemisphere equal-area projections; contours at 0.1 km s⁻¹ intervals for compressional waves, 0.02 km s⁻¹ intervals for shear waves, and 0.5% anisotropy intervals for shear wave polarization anisotropy. Dashed line marks minimum contours. Maximum P-wave and S-wave anisotropy is 0.9% and 1.9%, respectively. Black and white lines (for low and high anisotropy, respectively) in right panel indicate the direction of polarization of the fast shear wave.

weakly anisotropic upper mantle layer below 250 km depth is actively deforming by dislocation creep, and hence the top and bottom layers may be strongly coupled down to 400 km depth. Our predictions (that weak seismic anisotropy will develop in olivine-rich aggregates deforming by [001] (*hk*0) slip in the deep upper mantle) challenge the two traditional interpretations for regions in the deep Earth of weak seismic anisotropy; (1) that they represent zones of poor deformation coherence at the seismic length scale, or (2) that the dominant deformation mechanism (for example, diffusion creep) in these regions does not produce CPO. Indeed, transition from dominant [100] to [001] slip at high pressure may explain the variation with depth of the anisotropy patterns of P and S waves, even if the entire upper mantle deforms coherently with a dominant horizontal shearing component (as expected in a convective system with large-scale plates at the surface, like the Earth's mantle). □

Received 11 August; accepted 30 November 2004; doi:10.1038/nature03266.

- Montagner, J.-P. Seismic anisotropy of the Pacific Ocean inferred from long-period surface waves dispersion. *Phys. Earth Planet. Inter.* **38**, 28–50 (1985).
- Cara, M. & Lévêque, J. L. Anisotropy of the asthenosphere: The higher mode data of the Pacific revisited. *Geophys. Res. Lett.* **15**, 205–208 (1988).
- Nishimura, C. E. & Forsyth, D. W. The anisotropic structure of the upper mantle in the Pacific. *Geophys. J.* **96**, 203–230 (1989).
- Montagner, J.-P. & Kennett, B. L. N. How to reconcile body-wave and normal-mode reference earth models. *Geophys. J. Int.* **125**, 229–248 (1996).
- Karato, S.-I. On the Lehman discontinuity. *Geophys. Res. Lett.* **19**, 2255–2258 (1992).
- Karato, S. I. & Rubie, D. C. Toward an experimental study of deep mantle rheology: a new multianvil sample assembly for deformation studies under high pressures and temperatures. *J. Geophys. Res.* **102**, 20111–20122 (1997).
- Cordier, P. & Rubie, D. C. Plastic deformation of minerals under extreme pressure using a multi-anvil apparatus. *Mater. Sci. Eng. A* **309–310**, 38–43 (2001).
- Durham, W. B., Weidner, D. J., Karato, S. I. & Wang, Y. in *Plastic Deformation of Minerals and Rocks* (eds Karato, S.-I. & Wenk, H.-R.) (American Mineralogical Society, Washington, 2002).
- Couvy, H. *et al.* Shear deformation experiments of forsterite at 11 GPa – 1400 °C in the multianvil apparatus. *Eur. J. Mineral.* (in the press).
- Tommasi, A. Forward modeling of the development of seismic anisotropy in the upper mantle. *Earth Planet. Sci. Lett.* **160**, 1–13 (1998).
- Silver, P. G. Seismic anisotropy beneath the continents: Probing the depths of geology. *Annu. Rev. Earth Planet. Sci.* **24**, 385–432 (1996).
- Ekström, G. & Diwonski, A. M. The unique anisotropy of the Pacific upper mantle. *Nature* **394**, 168–172 (1998).
- Gung, Y., Panning, M. & Romanowicz, B. Global anisotropy and the thickness of continents. *Nature* **422**, 707–711 (2003).
- Lévêque, J. J., Debayle, E. & Maupin, V. Anisotropy in the Indian Ocean upper mantle from Rayleigh and Love waveform inversion. *Geophys. J. Int.* **133**, 529–540 (1998).
- Gaherty, J. B., Jordan, T. H. & Gee, L. S. Seismic structure of the upper mantle in the central Pacific corridor. *J. Geophys. Res.* **101**, 22291–22309 (1996).
- Revenaugh, J. & Jordan, T. Mantle layering from ScS reverberations 3. The upper mantle. *J. Geophys. Res.* **96**, 19781–19810 (1991).
- Li, L., Raterron, P., Weidner, D., Chen, J. & Vaughan, M. T. Stress measurements of deforming olivine at high pressure. *Phys. Earth Planet. Inter.* **143–144**, 357–367 (2004).
- Li, L., Raterron, P., Weidner, D. & Chen, J. Olivine flow mechanisms at 8 GPa. *Phys. Earth Planet. Inter.* **138**, 113–129 (2003).
- Ben Ismail, W. & Mainprice, D. An olivine fabric database: an overview of upper mantle fabrics and seismic anisotropy. *Tectonophysics* **296**, 145–158 (1998).
- Ben Ismail, W., Barruol, G. & Mainprice, D. The Kaapvaal craton seismic anisotropy: petrophysical analyses of upper mantle kimberlite nodules. *Geophys. Res. Lett.* **28**, 2497–2500 (2001).
- Vaucher, A., Dineur, F. & Rudnick, R. Microstructure, texture, and seismic anisotropy of the lithospheric mantle above a plume. Insights from the Labait volcano xenoliths. *Earth Planet. Sci. Lett.* (in the press).
- Lebensohn, R. A. & Tomé, C. N. A self-consistent anisotropic approach for the simulation of plastic deformation and texture development of polycrystals: Application to zirconium alloys. *Acta Metall. Mater.* **41**, 2611–2624 (1993).
- Wenk, H.-R., Bennet, K., Canova, G. R. & Molinari, A. Modelling plastic deformation of peridotite with the self-consistent theory. *J. Geophys. Res.* **96**, 8337–8349 (1991).
- Tommasi, A., Mainprice, D., Canova, G. & Chastel, Y. Viscoplastic self-consistent and equilibrium-based modeling of olivine lattice preferred orientations. Implications for upper mantle seismic anisotropy. *J. Geophys. Res.* **105**, 7893–7908 (2000).
- Abramson, E. H., Brown, M., Slutsky, L. J. & Zang, J. The elastic constants of San Carlos olivine up to 17 GPa. *J. Geophys. Res.* **102**, 12252–12263 (1997).
- Chai, M., Brown, J. M. & Slutsky, L. J. The elastic constants of a pyrope-grossular-almandine garnet up to 20 GPa. *Geophys. Res. Lett.* **24**, 523–526 (1997).
- Collins, M. D. & Brown, J. M. Elasticity of an upper mantle clinopyroxene. *Phys. Chem. Miner.* **26**, 7–13 (1998).
- Mainprice, D., Bascou, J., Cordier, P. & Tommasi, A. Crystal preferred orientations of garnet: Comparison between numerical simulations and electron back-scattered diffraction (EBSD) measurements in naturally deformed eclogites. *J. Struct. Geol.* **26**, 2089–2102 (2004).
- Bascou, J., Tommasi, A. & Mainprice, D. Plastic deformation and development of clinopyroxene lattice preferred orientations in eclogites. *J. Struct. Geol.* **24**, 1357–1368 (2002).

- Mainprice, D., Barruol, G. & Ben Ismail, W. in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale* (eds Karato, S.-I., Forte, A. M., Liebermann, R. C., Masters, G. & Stixrude, L.) 237–264 (AGU, Washington DC, 2000).

Acknowledgements This Letter is dedicated to the memory of G. Canova, who introduced A.T. and D.M. to VPSC modelling. H.C. was supported by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M. (David.Mainprice@dstu.univ-montp2.fr).

Stratigraphic placement and age of modern humans from Kibish, Ethiopia

Ian McDougall¹, Francis H. Brown² & John G. Fleagle³

¹Research School of Earth Sciences, Australian National University, Canberra, ACT 0200, Australia

²Department of Geology and Geophysics, University of Utah, Salt Lake City, Utah 84112, USA

³Department of Anatomical Science, Stony Brook University, Stony Brook, New York 11794, USA

In 1967 the Kibish Formation in southern Ethiopia yielded hominid cranial remains identified as early anatomically modern humans, assigned to *Homo sapiens*^{1–4}. However, the provenance and age of the fossils have been much debated^{5,6}. Here we confirm that the Omo I and Omo II hominid fossils are from similar stratigraphic levels in Member I of the Kibish Formation, despite the view that Omo I is more modern in appearance than Omo II^{1–3}. ⁴⁰Ar/³⁹Ar ages on feldspar crystals from pumice clasts within a tuff in Member I below the hominid levels place an older limit of 198 ± 14 kyr (weighted mean age 196 ± 2 kyr) on the hominids. A younger age limit of 104 ± 7 kyr is provided by feldspars from pumice clasts in a Member III tuff. Geological evidence indicates rapid deposition of each member of the Kibish Formation. Isotopic ages on the Kibish Formation correspond to ages of Mediterranean sapropels, which reflect increased flow of the Nile River, and necessarily increased flow of the Omo River. Thus the ⁴⁰Ar/³⁹Ar age measurements, together with the sapropel correlations, indicate that the hominid fossils have an age close to the older limit. Our preferred estimate of the age of the Kibish hominids is 195 ± 5 kyr, making them the earliest well-dated anatomically modern humans yet described.

The principal outcrops of the Kibish Formation are along the Omo River where it skirts the Nkalabong Range (Fig. 1), with the highest outcrops close in elevation to that of the watershed between the Omo and the Nile rivers^{7,8}. Former hydrographic links are apparent from Nilotic fauna in the Turkana Basin sequence^{8–10}.

The Kibish Formation (about 100 m thick) consists of flat-lying, tectonically undisturbed, unconsolidated sediments deposited mainly in deltaic environments over brief periods. It comprises the youngest exposed sedimentary sequence in the Omo Basin, and lies disconformably upon the Nkalabong Formation^{11,12} or on the underlying Mursi Formation¹². Strata are composed principally of claystone and siltstone, with subordinate fine sandstone, conglomerate and tuffs (Fig. 2).

Butzer *et al.*¹³ and Butzer¹⁴ divided the Kibish Formation into Members I to IV on the basis of disconformities with up to 30 m relief (Fig. 2). The members record discrete times of deposition

when the northern margin of Lake Turkana and the Omo delta lay about 100 km north of their current positions. As the higher lake levels reflect significantly higher precipitation in the region, such periods should be recognizable at least regionally, and perhaps globally.

Member I (26 m thick; Fig. 2) was deposited unconformably on the Nkalabong Formation in a deltaic environment. Small (less than 30 mm) rounded pumice clasts occur in an impure tuff in Member I, 7 m below the base of Member II. This tuff lies near, but probably slightly below, the levels from which Omo I and Omo II were derived. Glass shards from the tuff are very similar in composition to the glass from three pumice clasts enclosed in the tuff (Supplementary Table 1), indicating a common origin. Further, this composition is distinct from the glass composition of any other tuff in the Kibish Formation. Member II, about 28 m thick, was deposited on a topographic surface developed on Member I with at least 19 m of relief. Member II contains two discrete sequences, separated by an internal disconformity, designated Members IIa and IIb on Fig. 2. Member II was incised by as much as 25 m before the deposition of Member III^{1,14}, which begins with thin siltstone and claystone beds that drape topography. These beds, averaging about 3 cm thick, fine upward internally, and represent annual flooding. Thus, deposition of the lower part of Member III may record less than 1 kyr. A tuff 3.5–12 m thick, 18 m above the base

of Member III, locally contains pumice clasts with alkali feldspar phenocrysts. Glass of the pumices normally differs compositionally from the glass of the tuff (Supplementary Table 1). However, one pumice sample yielded two contrasting sets of analyses: one corresponding to the tuff, the other to the dominant pumices. The youngest unit of the Kibish Formation, Member IV, was deposited on the underlying sediments after up to 30 m of dissection. Member IV comprises at most 21 m of strata deposited between about 9.5 and 3.3 kyr ago, on the basis of ¹⁴C dating of mollusc shell^{7,14,15}.

Omo I was found at Kamoya's Hominid Site (KHS; 5° 24.15' N, 35° 55.81' E), which was identified from contemporary photographs, from evidence of the 1967 excavations and by additional hominid bone material conjoining the 1967 finds¹⁶. The hominid fossils were recovered from a siltstone 2.4 m below the base of Member II¹⁴. Thus, Omo I derives from near the top of Member I. Butzer *et al.*¹³ reported a ²³⁰Th/²³⁴U date of 130 ± 5 kyr on *Etheria*

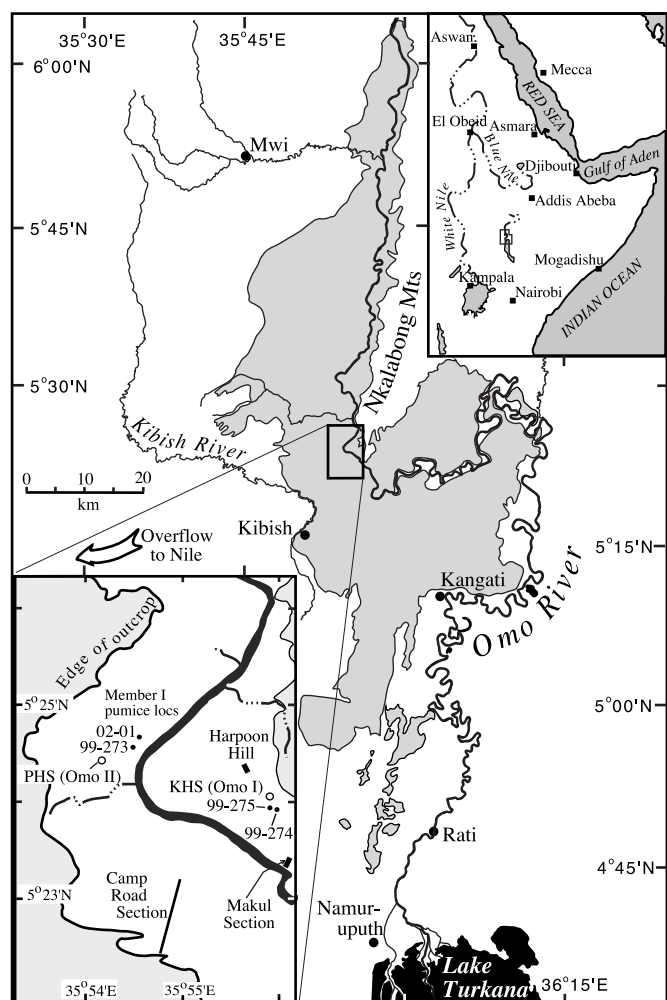


Figure 1 Map showing the distribution of the Kibish Formation (shaded) in the lower Omo Valley, southern Ethiopia, after Davidson²⁹. Inset on lower left, locations of Omo I, Omo II, measured sections, and dated samples.

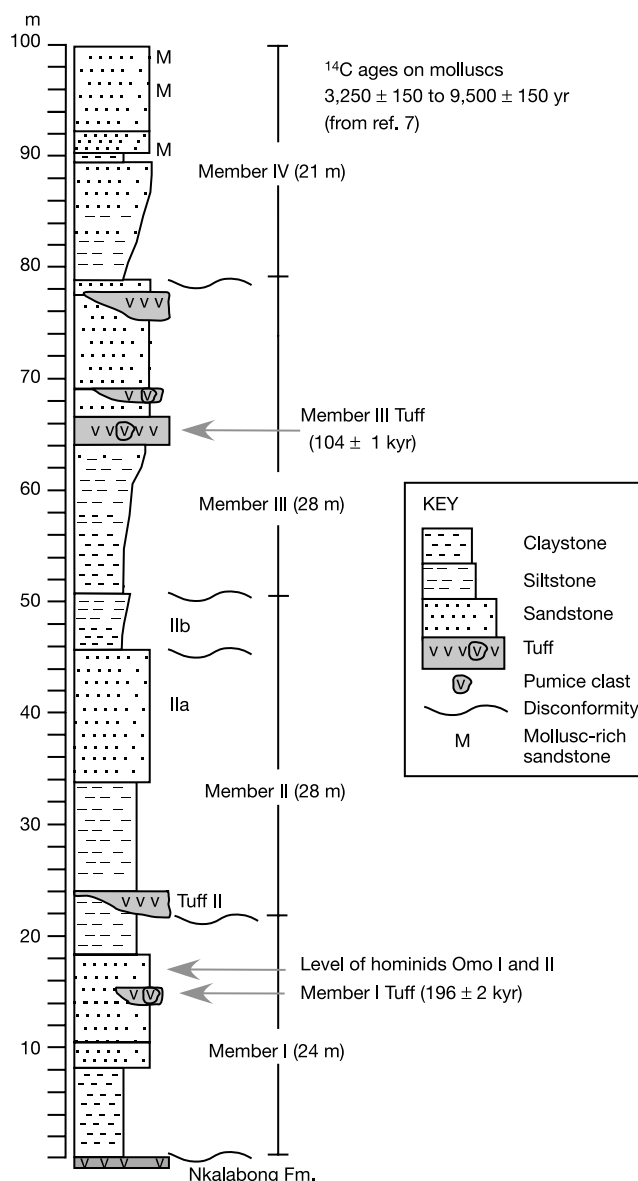


Figure 2 Composite stratigraphy of the Kibish Formation. Member I was measured near the type section at Makul; Member II was measured near Harpoon Hill, and Members III and IV were measured along Camp Road (see Fig. 1 for locations).

from essentially the same level as the hominid, but they and others have questioned the reliability of the age^{5,6}.

Omo II was found on the surface at PHS (Paul's Hominid Site), which Butzer¹⁴ mapped about 2.6 km northwest of KHS. A map drawn at the time by Paul Abell (discoverer of Omo II), together with contemporary photographs of the site supplied by K.W. Butzer, have positively identified PHS at 5° 24.55' N, 35° 54.07' E, about 3.3 km west by north of KHS (Fig. 1). Although PHS was mislocated on the published map, Butzer's¹⁴ stratigraphic description at the site is correct. The base of Member II lies about 3 m above the approximate level from which Omo II was recovered. The basal tuff of Member II is of unique composition, and correlates from KHS to PHS (Supplementary Table 1).

Ages on alkali feldspars separated from pumice clasts from tuffs in Members I and III reflect the time of their eruption, and provide maximum and minimum ages for the hominid fossils, respectively. ⁴⁰Ar/³⁹Ar results are summarized in Table 1, with details listed in Supplementary Tables 2–4.

The five pumice clasts measured from Member I yielded three different ages (Table 1), with at least the oldest age interpreted as evidence for reworking of that particular pumice clast. Nine analyses on 99-273A gave a mean age of 320 ± 19 kyr. Seven analyses on 99-273B yielded a mean age of 221 ± 13 kyr after rejecting one outlier (269 ± 18 kyr). Single feldspar crystals from 02-01B yielded a mean age of 195 ± 11 kyr (*n* = 14), forming a concordant data set after one rejection (333 ± 6 kyr). Multiple crystals (*n* ≤ 3) were analysed in 12 of 16 measurements for 02-01A, and in 8 of 14 analyses for 02-01C (*n* ≤ 4). No ages were rejected on 02-01A, whose mean age is 204 ± 15 kyr, or on 02-01C, whose mean age is 196 ± 15 kyr. It is probable that these latter three pumice clasts are products of the same volcanic eruption, as shown by their concordant ages and similar mean K/Ca ratios. When combined, all 44 analyses provide an arithmetic mean age of 198 ± 14 kyr and a weighted mean age of 195.8 ± 1.6 kyr. The mean age calculated for pumice 99-273B is statistically older than the ages determined on the three pumice clasts 02-01, so that 99-273B might also be a reworked pumice clast. Clearly, the age of deposition of a tuff must be younger than the youngest igneous component found within it. Thus, this tuffaceous level of Member I of the Kibish Formation was deposited after 196 kyr ago, on the basis of the mean age determined on the three 02-01 pumice clasts.

Feldspars from four pumice clasts from the Member III tuff were analysed. Two were *in situ* (99-275A, C), and two (99-274A, B) had weathered out of the same unit about 150 m farther east (Fig. 1). Ages on 99-275A range from 98.0 ± 7.7 kyr to 114.6 ± 4.3 kyr. Five single-crystal measurements have an arithmetic mean age of 105.5 ± 7.0 kyr, identical to five measurements on groups of two or three crystals (mean age 105.4 ± 2.7 kyr), indicating a homo-

geneous population. The overall mean age is 105.4 ± 5.0 kyr, with a weighted mean age of 106.0 ± 1.6 kyr. Eleven analyses on sample 99-275C yield a mean age of 107.5 ± 7.1 kyr and a weighted mean of 105.4 ± 1.8 kyr, after rejection of two outliers (142.1 ± 3.3 and 128.8 ± 8.3 kyr). Concordant results from 99-274A on seven single crystals and six pairs of crystals gave an overall arithmetic mean age of 98.1 ± 5.3 kyr. The companion clast 99-274B gave a mean age of 105.0 ± 8.3 kyr, after elimination of one outlier (Supplementary Table 4). Again these results reflect a single age population, as shown by the similar mean ages of the individual pumices, and the overlapping average K/Ca ratios. Combining all these results (except outliers), the overall arithmetic mean age is 103.7 ± 7.4 kyr (*n* = 47) and the weighted mean age is 103.7 ± 0.9 kyr. Thus, the depositional age must be equal to or younger than 104 kyr, providing evidence that Members I and II of the Kibish Formation are older than 104 kyr.

Each of the members of the Kibish Formation was deposited during intervals when Lake Turkana was at a much higher level than at present, and Member II has an internal unconformity. The upper part of Member I was being deposited at or after 196 kyr ago, and the upper part of Member III was being deposited at or after 104 kyr ago; ¹⁴C ages on Member IV correspond to deposition between 9.5 and 3.3 kyr ago. These ages are remarkably similar to ages of Mediterranean sapropels S7, S4 and S1. Sapropels S1–S7 have the following estimated ages: 195 kyr (S7), 172 kyr (S6), 124 kyr (S5), 102 kyr (S4), 81 kyr (S3), 55 kyr (S2) and 8 kyr (S1)¹⁷. In many cases sapropels are related to a greatly increased flow of the Nile River into the Mediterranean Sea as a consequence of intensification of the African monsoon¹⁸, recorded in more negative δ¹⁸O in planktonic foraminifera. As the Omo River shares a divide with the Blue Nile and with tributaries of the White Nile, the Nile and the Omo must be affected similarly. As noted, deposition of each of the members of the Kibish Formation was probably very rapid. Thus, the close correspondence between the ages of Member I (196 kyr) and of sapropel S7 (195 kyr), of Member III (104 kyr) and of sapropel S4 (102 kyr), and of Member IV (3.3–9.5 kyr) and of sapropel S1 (8 kyr) is probably causally related. Sapropel S2 (55 kyr) is absent from or poorly represented in many Mediterranean sedimentary cores and has a very small δ¹⁸O residual; thus, it is not surprising that no deposits of this age have been identified in the Kibish Formation. Sapropel S6, deposited during a European glacial period¹⁹, might also be absent from the Kibish Formation, because it too has a small δ¹⁸O anomaly. The two parts of Member II may be accommodated by sapropels S5 and S6, or they may correspond to the two phases identified in sapropel S5 (119–124 kyr), which are separated by 700–900 yr (ref. 20). This link between sapropel formation in the Mediterranean and very high levels of Lake Turkana is a particularly notable finding. In contrast, S3 is very well represented in many Mediterranean sedimentary cores and is

Table 1 Summary of ⁴⁰Ar/³⁹Ar alkali feldspar laser fusion ages from pumice clasts in tuffs of the Kibish area, Turkana Basin, Ethiopia

Sample no.	Tuff	Locality	Irradiation	<i>n</i>	<i>n</i> used	Simple mean age (kyr)	Weighted mean age (kyr)	Isochron age (kyr)	MSWD	(⁴⁰ Ar/ ³⁹ Ar) _i
Kibish Formation, Member III										
99-275A	Member III	0.4 km SSE of KHS	ANU58/L10	10	10	105.4 ± 5.0	106.0 ± 1.6	108.0 ± 3.1	1.03	292.2 ± 4.1
99-275C	Member III	0.4 km SSE of KHS	ANU58/L1	13	11	107.5 ± 7.1	105.4 ± 1.8	101.1 ± 8.4	2.04	303.6 ± 13.8
99-274A	Member III	0.5 km SE of KHS	ANU58/L6	13	13	98.1 ± 5.3	98.2 ± 1.1	97.4 ± 2.2	2.47	298.9 ± 6.8
99-274B	Member III	0.5 km SE of KHS	ANU58/L7	14	13	105.0 ± 8.3	109.6 ± 1.3	110.2 ± 4.4	3.27	292.9 ± 16.2
Kibish Formation, Member I, near Omo II site, just west of Omo River, Nakaa'kire, 5° 24.6' N, 35° 54.5' E										
99-273A	Member I	Nakaa'kire	ANU58/L3	9	9	319.8 ± 18.6	315.3 ± 3.6	318.2 ± 16.6	3.55	287.3 ± 41.5
99-273B	Member I	Nakaa'kire	ANU58/L4	8	7	220.8 ± 13.3	210.9 ± 2.1	209.6 ± 2.0	0.55	301.4 ± 3.1
02-01A	Member I	Nakaa'kire	ANU98/L9	16	16	204.2 ± 14.9	205.3 ± 2.2	194.1 ± 5.4	2.66	315.8 ± 7.3
02-01B	Member I	Nakaa'kire	ANU98/L10	15	14	194.7 ± 10.8	191.5 ± 1.9	184.9 ± 3.8	2.13	304.0 ± 3.5
02-01C	Member I	Nakaa'kire	ANU98/L3	14	14	195.6 ± 15.0	192.7 ± 2.3	192.5 ± 6.7	3.85	295.7 ± 5.8

⁴⁰K decay constant λ = 5.543 × 10^{−10} yr^{−1}. Fluence monitor: Fish Canyon Tuff sanidine 92-176 of reference age 28.1 Myr. Samples 99-275A and 99-275C (laboratory sample numbers at ANU) = KIB99-47 (field sample number); 99-274A and 99-274B = KIB99-41; 99-273A and 99-273B = KIB99-19. Results with errors are means ± s.d. MSWD, mean square of weighted deviates.

therefore expected to be recorded in the Kibish Formation, but has not been recognized. Given the large expanse of the plain in the region underlain by the Kibish Formation, it is quite possible that deposits correlative with sapropel S3 are present but are not exposed in the immediate Kibish region that we have studied.

Our palaeontological and stratigraphic studies support the original report¹⁴ that Omo I and Omo II are derived from comparable stratigraphic levels within Member I of the Kibish Formation despite their morphological differences^{1–3,21,22}. Morphological diversity among fossil hominids from the Middle and Late Pleistocene of Africa is of major importance in understanding the tempo and mode of modern human origins^{21,23}.

⁴⁰Ar/³⁹Ar dating of feldspars from tuffs in Member I and Member III of the Kibish Formation shows that its hominid fossils are younger than 195.8 ± 1.6 kyr and older than 103.7 ± 0.9 kyr. Direct isotopic dating of volcanic eruptions recorded in the Kibish Formation does not enable us to place narrower limits on the age. However, the suggested correlations of Member IIa and Member IIb with either the two identified phases of sapropel S5 or sapropels S6 and S5, respectively, indicate that deposition of Member I of the Kibish Formation occurred earlier than about 125 kyr ago or earlier than 172 kyr ago. The geological evidence for rapid deposition of Member I and the remarkably close correspondence of the isotopic ages on the youngest pumice clasts in the tuff of Member I at 196 kyr with the estimated age of sapropel S7 is regarded as strongly supporting the view that Member I was deposited close to 196 ± 2 kyr ago. On this basis we suggest that hominid fossils Omo I and Omo II are relatively securely dated to 195 ± 5 kyr old, somewhat older than the age of between 154 and 160 kyr assigned to the hominid fossils from Herto, Ethiopia²⁴, making Omo I and Omo II the oldest anatomically modern human fossils yet recovered. □

Methods

Age measurements

Alkali feldspar crystals were separated from pumice clasts and cleaned ultrasonically in 7% HF for 5–10 min to remove adhering volcanic glass and surface alteration. Although separations were performed on the coarsest crystals present, in several cases crystals were less than 1 mm, with masses less than 0.8 mg. In such cases several crystals were used for each analysis (see Supplementary Tables 2–4).

Samples were irradiated in facilities X33 or X34 of the High Flux Australian Reactor (Lucas Heights, Sydney, Australia) for 6 h as described in ref. 25. Cadmium shielding 0.2 mm thick was used to reduce the (⁴⁰Ar/³⁹Ar)_K correction factor, which was measured by analysis of zero-aged synthetic potassium silicate glass co-irradiated with the unknowns. This correction is particularly important because of the young age of the samples, so interpolated values for each sample are given in the Supplementary Tables 2–4. The fluence monitor employed was sanidine 92-176, separated from the Fish Canyon Tuff, with a reference age of 28.1 Myr (ref. 26).

After irradiation, feldspar crystals were loaded into wells in a copper sample tray, installed in the vacuum system and baked overnight. Samples were fused with a focused argon-ion laser beam with up to 10 W of power. After purification of the gases released during fusion, the argon was analysed isotopically in a VG3600 mass spectrometer, using a Daly collector. The overall sensitivity of the system was about 2.5×10^{-17} mol mV⁻¹. Mass discrimination was monitored through regular measurements of atmospheric argon. The irradiation parameter, *J*, for each unknown was derived by interpolation from the measurements made on the fluence monitor crystals, with at least five analyses per level; the precision generally was in the range 0.3–0.75%, standard deviation of the population. Calcium correction factors²⁷ used in all calculations were (³⁶Ar/³⁷Ar)_{Ca} = 3.49×10^{-4} and (³⁹Ar/³⁷Ar)_{Ca} = 7.86×10^{-4} .

Data handling

In calculating the arithmetic mean age for each pumice clast, any result more than two standard deviations from the initial mean of each sample was rejected iteratively until no further outliers were identified. No more than two measurements were rejected in any group of analyses, and no results were rejected for about half the groups. The error quoted in Table 1 is the standard deviation of the population, but because uncertainties on individual ages are variable, a weighted mean age and error is also given, weighting each age by the inverse of the variance. Differences in these mean ages are small (Table 1), but the error of the weighted mean age is usually much lower. In addition, data from each pumice clast, after exclusion of outliers, were plotted in an isotope correlation diagram (³⁶Ar/⁴⁰Ar versus ³⁹Ar/⁴⁰Ar), using the York²⁸ procedure. The derived ages are quite close to the mean ages (Table 1), and the calculated trapped argon composition generally has the atmospheric argon ratio of 295.5, within uncertainty.

Received 22 September; accepted 8 December 2004; doi:10.1038/nature03258.

- Day, M. H. Omo human skeletal remains. *Nature* **222**, 1135–1138 (1969).
- Day, M. H. & Stringer, C. B. in *Congrès International de Paléontologie Humaine I*, Nice Vol. 2, 814–846 (Colloque International du CNRS, 1982).
- Day, M. H. & Stringer, C. B. Les restes crâniens d'Omo-Kibish et leur classification à l'intérieur du genre *Homo*. *Anthropologie* **95**, 573–594 (1991).
- Day, M. H., Twist, M. H. C. & Ward, S. Les vestiges post-crâniens d'Omo I (Kibish). *Anthropologie* **95**, 595–610 (1991).
- Howell, F. C. in *Evolution of African Mammals* (eds Maglio, V. J. & Cooke, H. B. S.) 154–248 (Harvard Univ. Press, Cambridge, Massachusetts, 1978).
- Smith, F. H., Falsetti, A. B. & Donnelly, S. M. Modern human origins. *Yb Phys. Anthropol.* **32**, 35–68 (1989).
- Butzer, K. W., Isaac, G. L., Richardson, J. L. & Washbourn-Kamau, C. Radiocarbon dating of East African lake levels. *Science* **175**, 1069–1076 (1972).
- Fuchs, V. E. The geological history of the Lake Rudolf Basin, Kenya Colony. *Phil. Trans. R. Soc. Lond. B* **229**, 219–274 (1939).
- Arambourg, C. *Mission Scientifique de l'Omo 1932–1933. Géologie–Anthropologie–Paléontologie* (Muséum National d'Histoire Naturelle, Paris, 1935–1947).
- Butzer, K. W. The Lower Omo Basin: Geology, fauna and hominids of Plio-Pleistocene formations. *Naturwissenschaften* **58**, 7–16 (1971).
- Butzer, K. W. in *Earliest Man and Environments in the Lake Rudolf Basin* (eds Coppens, Y., Howell, F. C., Isaac, G. L. & Leakey, R. E. F.) 12–23 (Univ. of Chicago Press, Chicago, 1976).
- Butzer, K. W. & Thurber, D. L. Some late Cenozoic sedimentary formations of the Lower Omo Basin. *Nature* **222**, 1138–1143 (1969).
- Butzer, K. W., Brown, F. H. & Thurber, D. L. Horizontal sediments of the lower Omo Valley: the Kibish Formation. *Quaternaria* **11**, 15–29 (1969).
- Butzer, K. W. Geological interpretation of two Pleistocene hominid sites in the Lower Omo Basin. *Nature* **222**, 1133–1135 (1969).
- Owen, R. B., Barthelme, J. W., Renaut, R. W. & Vincens, A. Palaeolimnology and archaeology of Holocene deposits north-east of Lake Turkana, Kenya. *Nature* **298**, 523–529 (1982).
- Fleagle, J. et al. The Omo I partial skeleton from the Kibish Formation. *Am. J. Phys. Anthropol. Suppl.* **36**, 95 (2003).
- Lourens, L. J. et al. Evaluation of the Plio-Pleistocene astronomical timescale. *Paleoceanography* **11**, 391–413 (1996).
- Rosignol-Strick, M., Nesteroff, W., Olive, P. & Vergnaud-Grazzini, C. After the deluge: Mediterranean stagnation and sapropel formation. *Nature* **295**, 105–110 (1982).
- Rosignol-Strick, M. & Paterne, M. A synthetic pollen record of the eastern Mediterranean sapropels of the last 1 Ma: implications for the time-scale and formation of sapropels. *Mar. Geol.* **153**, 221–237 (1999).
- Rohling, E. J. et al. African monsoon variability during the previous interglacial maximum. *Earth Planet. Sci. Lett.* **202**, 61–75 (2002).
- Haile-Selassie, Y., Asfaw, B. & White, T. D. Hominid cranial remains from Upper Pleistocene deposits at Aduma, Middle Awash, Ethiopia. *Am. J. Phys. Anthropol.* **123**, 1–10 (2004).
- Rightmire, G. P. in *The Origins of Modern Humans: A World Survey of the Fossil Evidence* (eds Smith, F. H. & Spencer, F.) 295–325 (Alan R. Liss, New York, 1984).
- Howell, F. C. in *Origins of Anatomically Modern Humans* (eds Nitecki, M. H. & Nitecki, D. V.) 253–319 (Plenum, New York, 1994).
- Clark, J. D. et al. Stratigraphic, chronological and behavioural contexts of Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423**, 747–752 (2003).
- McDougall, I. & Feibel, C. S. Numerical age control for the Miocene–Pliocene succession at Lothagam, a hominoid-bearing sequence in the northern Kenya Rift. *J. Geol. Soc. Lond.* **156**, 731–745 (1999).
- Spell, T. L. & McDougall, I. Characterization and calibration of ⁴⁰Ar/³⁹Ar dating standards. *Chem. Geol.* **198**, 189–211 (2003).
- Spell, T. L., McDougall, I. & Dougeris, A. P. The Cerro Toledo Rhyolite, Jemez Volcanic Field, New Mexico: ⁴⁰Ar/³⁹Ar geochronology of eruptions between two caldera-forming events. *Bull. Geol. Soc. Am.* **108**, 1549–1566 (1996).
- York, D. Least squares fitting of a straight line with correlated errors. *Earth Planet. Sci. Lett.* **5**, 320–324 (1969).
- Davidson, A. *The Omo River Project* (Bulletin 2, Ethiopian Institute of Geological Surveys, Addis Ababa, 1983).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Mya, R. Maier and X. Zhang for technical support for the geochronology; participants in the Kibish expeditions between 1999 and 2003, including Z. Assefa, J. Shea, S. Yirga, J. Trapani and especially C. Feibel, B. Passey and C. Fuller for their geological contributions; and R. Leakey, K. Butzer and especially P. Abell for providing us with information and documents about the 1967 expedition to the Kibish area. We thank the Government of Ethiopia, the Ministry of Youth, Sports and Culture, the Authority for Research and Conservation of Cultural Heritage, and the National Museum of Ethiopia for permission to study the Kibish Formation. Support was provided by the National Science Foundation, the Leakey Foundation, the National Geographic Society and the Australian National University. Neutron irradiations were facilitated by the Australian Institute of Nuclear Science and Engineering and the Australian Nuclear Science and Technology Organization.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.McD (ian.mcdougall@anu.edu.au).

Parasites and climate synchronize red grouse populations

Isabella M. Cattadori¹, Daniel T. Haydon² & Peter J. Hudson¹

¹Center for Infectious Diseases Dynamics, Mueller Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

²Division of Environmental and Evolutionary Biology, University of Glasgow, Glasgow G12 8QQ, UK

There is circumstantial evidence that correlated climatic conditions can drive animal populations into synchronous fluctuations in abundance^{1–5}. However, it is unclear whether climate directly affects the survival and fecundity of individuals, or indirectly, by influencing food and natural enemies. Here we propose that climate affects trophic interactions and could be an important mechanism for synchronizing spatially distributed populations. We show that in specific years the size of red grouse populations in northern England either increases or decreases in synchrony. In these years, widespread and correlated climatic conditions during May and July affect populations regionally and influence the density-dependent transmission of the gastrointestinal nematode *Trichostrongylus tenuis*, a parasite that reduces grouse fecundity⁶. This in turn forces grouse populations into synchrony. We conclude that specific climatic events may lead to outbreaks of infectious diseases or pests that may cause dramatic, synchronized changes in the abundance of their hosts.

Nonlinear interactions between climate and population size can influence not only temporal changes in abundance but can also lead to spatial synchrony at a range of scales^{1–5,7,8}. About 50 years ago, Moran⁹ proposed that uncoupled populations with identical linear density-dependent structure will asymptotically synchronize in phase under the influence of correlated environmental perturbations, and the correlation between populations will equal the correlation between the extrinsic factors. Since Moran suggested his theorem, researchers have looked for evidence of the ‘Moran effect’ across animal taxa and investigated how environmental noise interacts with nonlinear population dynamics, the ‘nonlinear Moran effect’^{2,8,10}. However, researchers have yet to identify the underlying process that generates correlated fluctuations between populations in any natural system. In populations in which fluctuations are driven by a trophic interaction, such as predator–prey, host–parasitoid or host–parasite interactions, the correlated climatic effects may operate either on these density-dependent mechanisms or more directly on population abundance, and thus move populations into a synchronous phase. Here we examine this hypothesis in the natural system of the red grouse (*Lagopus lagopus scoticus*) and its gastrointestinal parasite *Trichostrongylus tenuis*, a nematode known to play a major role in reducing fecundity and destabilizing grouse populations^{6,11,12}.

We used 91 time series of annual red grouse harvesting data obtained from managed grouse moors in northern England between 1839 and 1994 and collected by The Game Conservancy Trust (Fig. 1). Red grouse harvest data have proved to be a fair reflection of spatio-temporal changes in population density and annual productivity¹³. The majority (59%) of these time series exhibit cyclic fluctuations with an average period of 7 years (range 3–13 years)^{14,15}. Each grouse population is embedded within one of five discrete regions of contiguous habitat¹⁵ and we analysed synchrony between populations in each of these regions. Common statistical techniques can be used to identify spatial synchrony¹⁶ and provide an estimate of average synchrony between time series, but they fail to identify the years in which time series are forced into synchrony. To achieve this, the dynamical state of each red grouse population was represented in a state-based Markov chain model

that described temporal transitions between four phase states based on three consecutive years of observations: a cycle trough, a consecutive increase, a cycle peak and a consecutive decrease. For each region, we examined changes in the annual proportion of populations in each state, which allowed us to identify the years in which populations were forced into significant state-synchrony and quantify the state that these populations converge onto, in these years¹⁷. As an annual index of state synchrony among populations within each of the five regions, we applied the Shannon–Weaver diversity index to the proportion of populations in each state¹⁷. Populations were considered to be in synchrony when the diversity index deviated below the lower 95% confidence interval predicted by the Markov state transition model under the null hypothesis of no coupling between populations. Hence, a synchronous year can arise either as a dynamically predictable episode, as a consequence of synchrony in the preceding year, or as an unpredicted ‘collective forcing episode’ (CFE) that brings previously asynchronous populations into the same synchronous state^{8,10}. To identify years when these forcing events occurred, we calculated whether the diversity index fell outside the one-step 95% confidence interval, conditional on the state configuration in the previous year¹⁷. We focused our analyses on the years in which these forcing events occurred and looked for the putative synchronizing mechanism.

The diversity index calculated among populations within each of the five regions revealed that no regular pattern characterized the incidence of the forcing episodes (Fig. 2a). The majority of populations converged on a common dynamical state: ‘upward’ CFEs—either two successive years of increased abundance (32% of CFEs) or peaks (32% of CFEs)—or, less frequently, ‘downward’ CFEs—when populations converged on two successive years of population decline (8% of CFEs) or a trough (28% of CFEs) (Fig. 2b). On the basis of knowledge of the red grouse system, we propose that there are two major demographic circumstances that would tend to generate such collective forcing episodes in these harvesting data:

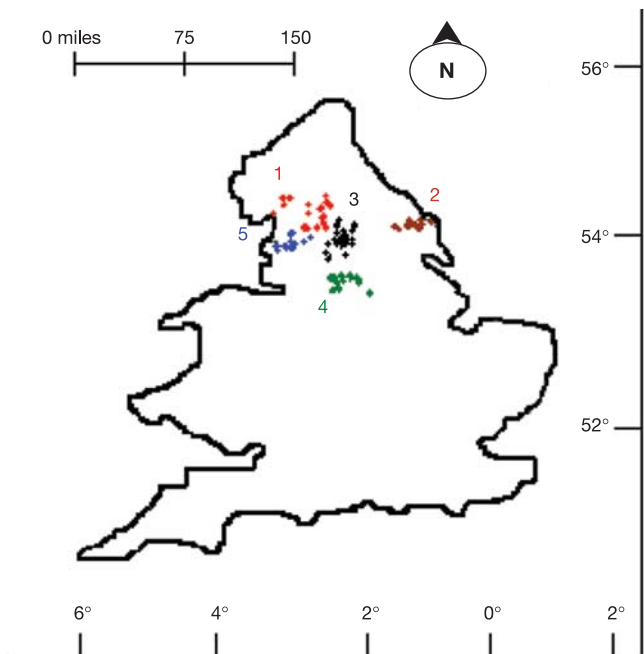


Figure 1 Locations of the 91 red grouse populations that provided annual harvesting records between 1839 and 1994. Populations were aggregated into five geographically distinct regions of contiguous grouse habitat (heather-dominant moorland), each represented here by a different colour. Synchrony between populations was examined within each of these distinct regions.

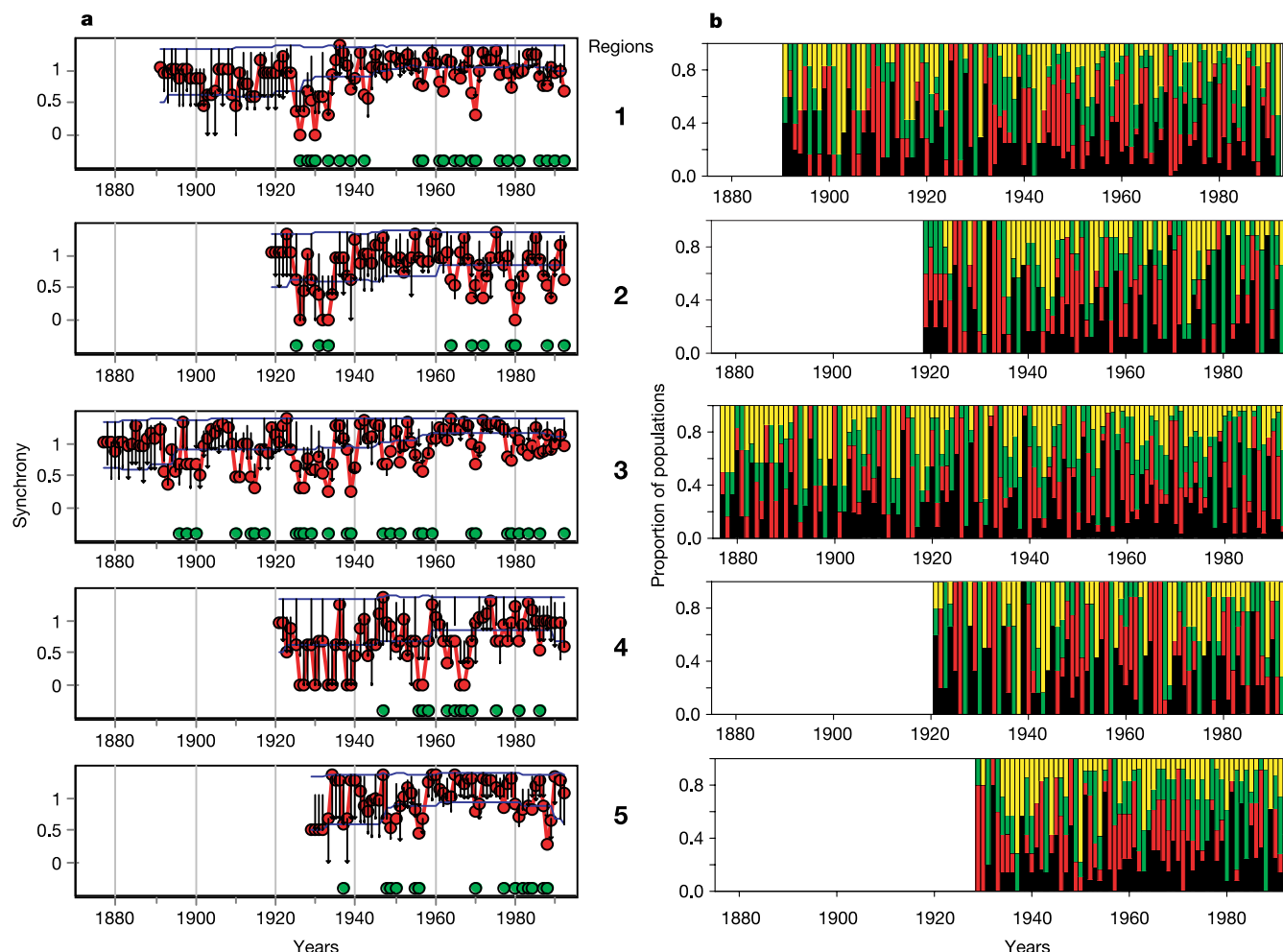


Figure 2 Spatio-temporal patterns of synchrony between populations within each of the five regions. **a**, The observed average state synchrony among populations (red line) expressed as the Shannon–Weaver index with the 95% confidence envelope (blue lines). Years in which CFEs forced the populations into an unexpected synchronous state are shown with green circles. These years were identified by predicting the expected average state of the populations conditional on the average state configurations observed the

previous year and estimating their 95 percentile interval (black bars)¹⁷. **b**, The proportion of populations observed in each state phase in each year. Troughs are shown in black, increases in red, peaks in green and decreases in yellow. Note that in collective forcing years the populations tend to enter the same state phase. For each region, phase synchrony and the proportion of phase states were calculated using sample sizes of at least five time series per year.

first, good breeding years in which populations undergo positive growth rates (increases or peaks) and second, poor breeding years when populations experience negative growth rates (decreases or troughs). Replicated field experiments have shown that the intensity of the nematode *T. tenuis* infection has a consistent and negative impact on female fecundity, clutch size, hatching success and chick survival^{11,18}. This parasite inhabits the blind-ending caeca of the grouse and has a direct life cycle. Eggs pass out of the host in the caecal faeces and in wet and warm conditions develop into a third-stage infective larva that ascends the heather and is ingested by a feeding grouse¹⁹. Temperature increases the development rate of the free-living stages and sufficient humidity is necessary for infection to be effective^{19–21}. Climatic conditions can also directly influence the survival of grouse chicks. Dry, warm weather in May, when chicks hatch, allows them to forage and grow, whereas chicks that hatch during cold, rainy periods have less foraging time and increased mortality²². Intrinsic processes such as changes in spacing behaviour late in the year may also influence grouse populations²³. However, experimental evidence has shown that such processes act at a local scale and are not associated with food and weather²⁴, and consequently are unlikely to be involved in generating the large-

scale intermittent collective forcing episodes and the pattern of synchrony that we have observed.

To identify the environmental variables that could force populations into synchrony, we applied a generalized logistic regression model which predicted the occurrence of annual forcing episodes as a function of the intensity of parasite infection and the climatic

Table 1 Environmental variables affecting CFEs in red grouse dynamics

	Coefficients $\pm$ standard error	Wald test	P
Total May rainfall (mm)	-0.11 ± 0.02	23.76	0.001
Total July rainfall (mm)	0.15 ± 0.04	12.85	0.001
Mean May temperature ($^{\circ}\text{C}$)	-0.83 ± 0.20	17.27	0.001
Mean July temperature ($^{\circ}\text{C}$)	0.88 ± 0.23	14.57	0.001
Mean parasite intensity (worm/host)	0.53 ± 0.20	7.17	0.001
Variables not in the model			
Total June rainfall (mm)	0.02 ± 0.02	1.49	0.22
Total August rainfall (mm)	-0.01 ± 0.01	0.08	0.78
Mean June temperature ($^{\circ}\text{C}$)	-0.06 ± 0.24	0.07	0.79
Mean August temperature ($^{\circ}\text{C}$)	-0.32 ± 0.19	2.93	0.09

Generalized logistic multiple regression model: backward stepwise. Degrees of freedom = 350; log-likelihood = -125.22 .

Table 2 Mean of environmental variables in synchronous 'upwards' and 'downwards' collective forcing years

	Upwards $\pm$ standard error ($n = 35$)	Downwards $\pm$ standard error ($n = 23$)	Mann-Whitney P
Total May rainfall (mm)	36.39 $\pm$ 1.21	47.75 $\pm$ 1.10	0.001
Total July rainfall (mm)	23.12 $\pm$ 0.81	19.47 $\pm$ 0.88	0.01
Mean May temperature ($^{\circ}\text{C}$)	10.72 $\pm$ 0.15	10.05 $\pm$ 0.15	0.01
Mean July temperature ($^{\circ}\text{C}$)	14.46 $\pm$ 0.16	15.95 $\pm$ 0.36	0.01
Parasite intensity (worm/host)	7.01 $\pm$ 0.14	7.50 $\pm$ 0.13	0.01

In 'upwards' collective forcing years, red grouse synchronize in peak or increase states; in 'downwards' collective forcing years, red grouse synchronize in decrease or trough states.

variables that have a major impact on breeding^{11,15,25}. This model indicated that parasite intensity and the rainfall and temperature during July promoted CFEs, while May rainfall and temperature repressed CFEs (Table 1). We examined two alternative hypotheses: first, that the direct effects of correlated climatic conditions on the grouse generate CFEs (the climate hypothesis), and second, that climate indirectly influences grouse dynamics through its effect on parasite transmission (the climate–parasite hypothesis).

A structural equation framework^{26,27}, based on the environmental variables selected by the logistic regression, was used to compare these hypotheses. For the climate hypothesis the model consisted of

rainfall and mean temperature in May and July with direct effects on grouse abundance, whereas for the climate–parasite hypothesis these climatic variables influenced the grouse both directly and indirectly through the net rate of parasite transmission, as measured by changes in parasite intensity in the host. These environmental variables differed consistently between the 'upward' and 'downward' synchronous forcing events such that Mays were dryer, Julys were wetter and parasite intensity was lower during the grouse 'upward' CFEs, whereas the pattern was reversed in the 'downward' CFEs (Table 2). Consequently, both models were fitted to these data. The two parsimonious models—based on the climate (Fig. 3a and b)

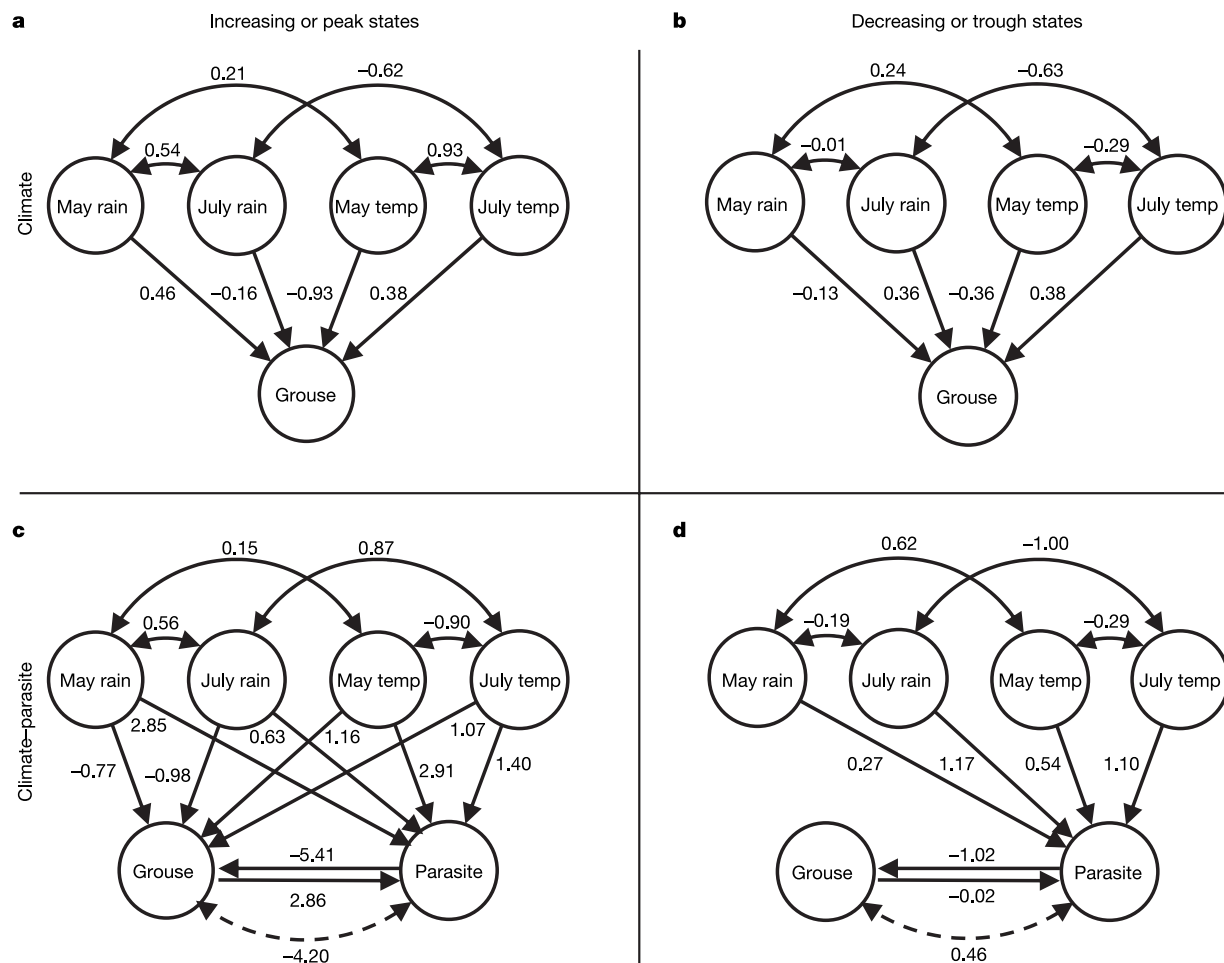


Figure 3 Simplified path diagram with parameterized path coefficients of simulated interactions between climate, parasites and red grouse. Explanatory variables describing each latent variable, the residual variance of each variable and the environmental noise of grouse and parasite variables are not represented²⁶. Circles, estimated latent variables. Curved double arrows, correlated variables; dotted and curved double arrows, variables affected by correlated environmental noise; single arrows, causal relationship between

variables. **a, c**, Based on data from years identified as collectively forcing grouse into synchronous increasing or peak states. **b, d**, Based on data from years identified as collectively forcing grouse into synchronous decreasing or trough states. The two parsimonious models are based on the red grouse climate hypothesis (**a, b**), and the red grouse climate–parasite hypothesis (**c, d**).

and climate–parasite (Fig. 3c and d) hypotheses—performed statistically well in both the ‘upward’ and ‘downward’ years, and there was no apparent statistical difference between them. For Fig. 3a, the generalized least squares (GLS) is 0.18, the probability that χ^2 of the observed hypothesis (H_{obs}) is bigger than the null hypothesis (H_{exp}), $P[\chi^2(H_{\text{obs}} > H_{\text{exp}})]$, is 0.83 and the corrected Akaike information criterion for small sample size (AIC_C) is 1.20. For Fig. 3c, $\text{GLS} = 2.03$, $P[\chi^2(H_{\text{obs}} > H_{\text{exp}})] = 0.86$ and $\text{AIC}_C = 1.55$. For Fig. 3b, $\text{GLS} = 4.76$, $P[\chi^2(H_{\text{obs}} > H_{\text{exp}})] = 0.96$ and $\text{AIC}_C = 2.00$. For Fig. 3d, $\text{GLS} = 7.33$, $P[\chi^2(H_{\text{obs}} > H_{\text{exp}})] = 0.99$ and $\text{AIC}_C = 3.53$.

However, a closer examination of the estimated grouse–environment interaction coefficients showed that the climate model was not ecologically consistent with the data in Table 2. Indeed, this model proposed that for the set of ‘upward’ CFEs (Fig. 3a) grouse abundance increased owing to wetter Mays and drier Julys, and that for the set of ‘downward’ CFEs (Fig. 3b) grouse abundance decreased owing to warmer Mays and colder Julys. This contrasts with both the data summarized in Table 2 and previous field studies^{15,25}. Thus the fit of the climate model is statistically reasonable but not biologically credible. The model based on the climate–parasite hypothesis was biologically consistent and in line with studies on the impact of weather both on grouse demography and on parasite transmission^{11,15,19}.

When applied to the set of ‘upward’ CFEs, the model revealed that relatively dry and warm Mays followed by cold and wet Julys were deleterious to parasite transmission, leading to low parasite intensities and positive red grouse growth rates. This collectively synchronized populations (Fig. 3c and Table 2). Within this model, climate also had a direct influence on grouse production, but the effect was less apparent. When applied to ‘downward’ CFEs, the climate–parasite model revealed that relatively wet Mays and warm Julys increased parasite transmission, which raised parasite intensity and led to a collective crash in grouse populations (Fig. 3d, Table 2). Dry and warm Mays impede the development of parasite eggs and the survival of the infective stage^{19–21}. If these climatic conditions are followed by cold wet Julys, parasite development is further reduced^{19–21}. Consequently, grouse feeding on the heather will have a low level of parasite infection, which results in an increase in grouse abundance and populations moving into a synchronous increase phase. In contrast, wet Mays and warm Julys favour abundant infective stages on the vegetation, leading to a rapid rise in the intensity of parasitic infection and causing a synchronous decline in grouse populations.

This study suggests that the effects of common climatic events on the density-dependent destabilizing interaction between host and parasite result in a phase synchrony between grouse populations. Our results also suggest that synchrony in grouse populations is an intermittent event, arising from populations that have been forced to converge into different phase states. These results have important implications for the management of natural populations and the control of infectious diseases because they indicate that certain weather conditions have a strong influence on eruptive outbreaks of infectious diseases or pest species and may lead to dramatic and synchronized changes in host abundance. □

Methods

Data sets

The red grouse time series provide one of the best data sets available on changes in relative abundance for examining spatio-temporal dynamics in animal populations. The time series are a good proxy for population abundance^{13,28}. Furthermore, detailed population studies on red grouse have been undertaken and the fundamental mechanisms of population regulation identified and tested experimentally^{6,11,25}. Each time series is based on a privately owned estate, the location of which was selected as the centre of the moorland habitat. For 60% of the time series we had corresponding time series of average intensity of *T. tenuis* infection in old and young red grouse randomly sampled between August and December for the period 1977–1994. *T. tenuis* intensity changes seasonally¹⁹ but reaches an autumnal asymptote²⁹. Monthly weather variables from April to August for

the period 1839–1994, namely millimetres of rainfall and mean temperature in degrees Celsius, were provided as daily time series by the UK Met Office–British Atmospheric Data Centre service (<http://badc.nerc.ac.uk/home>). Rainfall records were from weather stations less than 10 km from each moor, while temperature came from stations less than 15 km away. Temperature time series were poorly represented, and so each series was reconstructed by scaling the series to the temperature expected at sea level (assuming temperature changes of 6.4 °C every 100 m). For each region and each month the new mean-temperature time series was calculated and then rescaled to the altitude of each original time series. (see Supplementary Information).

Spatio-temporal synchrony

Red grouse time series and climatic time series had occasional missing data points (1.3% of the total cases in red grouse) that were interpolated with a log-linear fit¹⁴. The red grouse data were normalized using a Box–Cox transformation and the long-term trend removed by selecting the residuals from fitting a non-parametric LOESS function to the data (the smoothing parameter included 75% of the observations and gaussian errors were assumed), thus leaving the short-term pattern unchanged. Parasite time series were $\log(x + 1)$ -transformed and when fewer than three consecutive data points were missing they were replaced with the annual parasite mean from the local region. Significant collective forcing episodes were identified, and 95% confidence intervals were estimated from 5,000 bootstraps of the Shannon–Weaver index based on the state-based Markov chain model, using sample sizes of at least five time series per year¹⁷. We also carried out state-based Markov chain modelling for the parasite time series and the rainfall time series and found that the years of synchronous collective forcing in parasite time series coincided in 67% of the cases with synchronous CFEs for May rainfall and 50% of the cases with July rainfall (see Supplementary Information).

Causal mechanisms of synchrony

The structural equation model (SEM) was based on conceptual, inferential variables (known as latents) estimated through explanatory variables (known as manifests) represented by our time series. The SEM allowed us to explicitly account for environmental stochasticity and reciprocal relationships between variables, and as such correctly measures these latent variables and their interactions^{26,27}. The explanatory variables inferring the weather variables were assumed to be fixed. Parameter identification and their 95% confidence intervals were estimated using bootstrap and Monte Carlo techniques (SEPATH, StatSoft) because the data violated some of the assumptions of structural equation modelling²⁶. We first estimated the predicted model parameters and the χ -squared statistic for the generalized least-squares discrepancy function, using the correlation matrix of the original data set. We then carried out bootstrap extractions with replacement based on 50 samples and calculated the χ -squared statistic for the SEM simulated on these new data. Finally, we ran 1,000 Monte Carlo replications of these bootstrapped models and calculated the empirical probability $P(H_{\text{obs}} > H_{\text{exp}})$ and the 95% confidence intervals for the original model and its parameters (see Supplementary Information).

Received 13 July; accepted 15 December 2004; doi:10.1038/nature03276.

- Cattadori, I. M., Merler, S. & Hudson, P. J. Searching for mechanisms of synchrony in spatially structured gamebird populations. *J. Anim. Ecol.* **69**, 620–638 (2000).
- Grenfell, B. T. et al. Noise and determinism in synchronized sheep dynamics. *Nature* **394**, 674–677 (1998).
- Post, E. & Forchhammer, M. C. Synchronization of animal population dynamics by large-scale climate. *Nature* **420**, 168–171 (2002).
- Ranta, E., Kaitala, V., Lindström, J. & Lindén, H. Synchrony in population dynamics. *Proc. R. Soc. Lond. B* **262**, 113–118 (1995).
- Stenseth, N. C. et al. Common dynamic structure of Canada lynx populations within three climatic regions. *Science* **285**, 1071–1073 (1999).
- Hudson, P. J., Dobson, A. P. & Newborn, D. Population cycles and parasitism. *Science* **286**, 2425a (1999).
- Björnstad, O. N., Peltonen, M., Liebhold, A. M. & Baltensweiler, W. Waves of larch budmoth outbreaks in the European Alps. *Science* **298**, 1020–1023 (2002).
- Ranta, E., Kaitala, V., Lindström, J. & Helle, A. The Moran effect and synchrony in population dynamics. *Oikos* **78**, 136–142 (1997).
- Moran, P. A. P. The statistical analysis of the Canadian lynx cycle. II. Synchronization and meteorology. *Austr. J. Zool.* **1**, 291–298 (1953).
- Greenman, J. V. & Benton, T. G. The impact of stochasticity on the behaviour of nonlinear population models: synchrony and the Moran effect. *Oikos* **93**, 343–351 (2001).
- Hudson, P. J., Newborn, D. & Dobson, A. P. Regulation and stability of a free-living host–parasite system: *Trichostrongylus tenuis* in red grouse. I: Monitoring and parasite reduction experiments. *J. Anim. Ecol.* **61**, 477–486 (1992).
- Dobson, A. P. & Hudson, P. J. Regulation and stability of a free-living host–parasite system: *Trichostrongylus tenuis* in red grouse. II: Population models. *J. Anim. Ecol.* **61**, 487–498 (1992).
- Cattadori, I. M., Haydon, D. T., Thirgood, S. J. & Hudson, P. J. Are indirect measures of abundance a useful index of population density? The case of red grouse harvesting. *Oikos* **100**, 439–446 (2003).
- Haydon, D. T., Shaw, D. J., Cattadori, I. M., Hudson, P. J. & Thirgood, S. J. Analysing noisy time series: describing regional variation in the cyclic dynamics of red grouse. *Proc. R. Soc. Lond. B* **269**, 1609–1617 (2002).
- Hudson, P. J. *Grouse in Space and Time* (The Game Conservancy Trust, Fordingbridge, 1992).
- Buonaccorsi, J. P., Elkinton, J. S., Evans, S. R. & Liebhold, A. M. Measuring and testing for spatial synchrony. *Ecology* **82**, 1668–1679 (2001).
- Haydon, D. T., Greenwood, P. E., Stenseth, N. C. & Saitoh, T. Spatio-temporal dynamics of the grey-sided vole in Hokkaido: identifying coupling using state-based Markov-chain modelling. *Proc. R. Soc. Lond. B* **270**, 435–445 (2003).

18. Hudson, P. J. The effect of a parasitic nematode on the breeding production of red grouse. *J. Anim. Ecol.* **55**, 85–94 (1986).
19. Anderson, R. C. *Nematode Parasites of Vertebrates. Their Development and Transmission* 2nd edn (CABI, Wallingford, 2000).
20. Lee, D. L. *The Biology of Nematodes* (Taylor and Francis, London, 2002).
21. Marquardt, W. C., Demaree, R. S. & Grieve, R. B. *Parasitology and Vector Biology* (Harcourt Academic, San Diego, 2000).
22. Slagsvold, T. Production of young by the Willow Grouse *Lagopus lagopus* (L.) in Norway in relation to temperature. *Norw. J. Zool.* **23**, 269–275 (1975).
23. Mougeot, F., Redpath, S. M., Leckie, F. & Hudson, P. J. The effect of aggressiveness on the population dynamics of a territorial bird. *Nature* **421**, 737–739 (2003).
24. Watson, A., Moss, R. & Parr, R. Effect of food enrichment on numbers and spacing behaviour of red grouse. *J. Anim. Ecol.* **53**, 663–678 (1984).
25. Hudson, P. J. *et al.* Trophic interactions and population growth rates: describing patterns and identifying mechanisms. *Phil. Trans. R. Soc. Lond. B* **357**, 1259–1271 (2002).
26. Shipley, B. *Cause and Correlation in Biology* (Cambridge Univ. Press, Cambridge, 2000).
27. Pugesek, B. H., Tomer, A. & von Eye, A. *Structural Equation Modeling* (Cambridge Univ. Press, Cambridge, 2003).
28. Hudson, P. J., Dobson, A. P. & Newborn, D. in *Population Cycles* (ed. Berryman, A.) 109–129 (Oxford Univ. Press, Oxford, 2003).
29. Hudson, P. J. & Dobson, A. P. in *Infectious Diseases in Natural Populations* (eds Grenfell, B. T. & Dobson, A. P.) 144–176 (Cambridge Univ. Press, Cambridge, 1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements I.M.C. thanks D. Shaw for statistical advice. We are also grateful to D. Newborn of the Game Conservancy Trust, The Earl Peel and moor owners in northern England who allowed us to aggregate the database. O. Bjørnstad, E. Post and D. Johnson provided comments on an earlier version. I.M.C. was funded by a Marie Curie fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.M.C. (imc3@psu.edu).

Addition of human melanopsin renders mammalian cells photoresponsive

Z. Melyan¹, E. E. Tarttelin^{1,2}, J. Bellingham², R. J. Lucas² & M. W. Hankins¹

¹Department of Visual Neuroscience, Division of Neuroscience and Psychological Medicine, Imperial College London, Charing Cross Hospital Campus, Fulham Palace Road, London W6 8RF, UK

²Faculty of Life Sciences, Michael Smith Building, University of Manchester, Manchester M13 9PT, UK

A small number of mammalian retinal ganglion cells act as photoreceptors for regulating certain non-image forming photoresponses^{1–10}. These intrinsically photosensitive retinal ganglion cells express the putative photopigment melanopsin^{11–13}. Ablation of the melanopsin gene renders these cells insensitive to light¹⁴; however, the precise role of melanopsin in supporting cellular photosensitivity is unconfirmed. Here we show that heterologous expression of human melanopsin in a mouse paraneuronal cell line (Neuro-2a) is sufficient to render these cells photoreceptive. Under such conditions, melanopsin acts as a sensory photopigment, coupled to a native ion channel via a G-protein signalling cascade, to drive physiological light detection. The melanopsin photoresponse relies on the presence of *cis*-isomers of retinaldehyde and is selectively sensitive to short-wavelength light. We also present evidence to show that melanopsin functions as a bistable pigment in this system, having an intrinsic photoisomerase regeneration function that is chromatically shifted to longer wavelengths.

In order to determine the function of melanopsin in an intact

cellular environment, we examined the ability of heterologous expression of human melanopsin to render mammalian neuronal cells photoreceptive. Cells from the Neuro-2a mouse neuroblastoma cell line were transfected with plasmid expression vectors engineered to express both the complete coding sequence of human melanopsin under control of the immediate early promoter of cytomegalovirus (CMV), and an enhanced green fluorescent protein (EGFP) reporter gene driven by an SV40 enhancer/promoter. High transfection efficiencies were obtained using the lipofectamine technique, with up to 60% of Neuro-2a cells expressing EGFP as assessed by fluorescence microscopy. Untransfected Neuro-2a cells did not express melanopsin (or rod/cone opsins) in either their differentiated or undifferentiated states (see Supplementary Fig. 1). After plasmid transfection, human melanopsin was observed at both the mRNA and protein level (Fig. 1b; see Supplementary Figs 1 and 2).

Because opsin photopigments use *cis*-isomers of retinaldehyde as a chromophore, physiological light responses were first assessed following pre-incubation with commercially available 9-*cis*-retinaldehyde (1 h, 20 μ M). A 10 s stimulus of 420-nm light (20-nm half-bandwidth) resulted in a sustained inward current (recovering over >10 min), which was not observed in untransfected cells or cells transfected with a control vector lacking melanopsin (Fig. 1d, e). The magnitude of the response was irradiance-dependent (Fig. 1d), and under voltage ramps had a peak slope conductance for the brightest stimulus of 6.46 ± 0.82 nS (mean $\pm$ s.e.m., $n = 13$), some three times greater than previously observed in intrinsically photosensitive retinal ganglion cells (ipRGCs) following exposure to a stimulus of similar intensity⁵.

The photoreceptive function of melanopsin was dependent on the presence of an appropriate isoform of retinaldehyde (Fig. 2a). In the absence of exogenous retinaldehyde, melanopsin-transfected cells were not light-sensitive. Significant photosensitivity was observed following pre-incubation with either 9-*cis*- or 11-*cis*- but not with all-*trans*-retinaldehyde.

These findings are consistent with the hypothesis that melanopsin acts as a photopigment with a specific affinity for *cis*-isomers of retinaldehyde. However, application of all-*trans*-retinaldehyde in this preparation also seemed to support a modest light response (Fig. 2a). Phylogenetically, melanopsin is most closely related to the cephalopod rhodopsins¹⁵, which use 11-*cis*-retinaldehyde as a chromophore for sensory functions but are also capable of binding all-*trans*-retinaldehyde, which they re-isomerize to 11-*cis*-retinaldehyde upon appropriate light exposure¹⁶. Such an isomerase function would facilitate photosensitivity following incubation with all-*trans*-retinaldehyde provided that the test stimulus was of sufficient duration to generate a *cis*-isoform in the first instance.

To examine the possibility that melanopsin has a similar inherent chromophore regeneration capability, we assessed the ability of prior light exposure with a wavelength outside the response range of 9- or 11-*cis*-retinaldehyde-loaded cells (Fig. 3a) to facilitate photosensitivity of cells loaded with all-*trans*-retinaldehyde. Exposure to 540-nm light had no effect on membrane current, but responses of cells loaded with all-*trans*-retinaldehyde to a test stimulus of 420-nm light were greatly enhanced following exposure to the longer wavelength light (Fig. 2b–d). The hypothesis that this sensitization is caused by a photoisomerization of the all-*trans*-retinaldehyde to a *cis*-isomer is supported by the observation that 540-nm light had no effect on the responses of cells loaded with 11-*cis*-retinaldehyde. Interestingly, a small increase in photosensitivity was observed in 9-*cis*-retinaldehyde-loaded cells treated with 540-nm light. This would be predicted if photoisomerization resulted in the generation of 11-*cis*-retinaldehyde, as opsin photopigments incorporating this isoform are commonly more sensitive¹⁷ (Fig. 2a).

Neuro-2a cells do not express either of the known mammalian photoisomerases (RGR-opsin¹⁸ and peropsin), as determined by

reverse transcriptase polymerase chain reaction (RT-PCR, data not shown), suggesting that melanopsin itself acts as a photoisomerase in these cells. However, as addition of *cis*-retinaldehyde induced photosensitivity only in cells expressing melanopsin, such a function alone cannot explain the effects of human melanopsin expression in Neuro-2a cells. Rather, our findings support the hypothesis that melanopsin is a bistable pigment, employing *cis*-isoforms of retinaldehyde in its photosensory function and also acting as a photoisomerase to regenerate bleached chromophore. Such a photochemistry would have obvious advantages for a photopigment located in the inner retina, distant from the retinal pigment epithelium (the primary site of chromophore regeneration); indeed, it has been reported for a different vertebrate photopigment that occupies a similar environment¹⁹.

Although direct measures of spectral sensitivity in human ipRGCs are unavailable, action spectrum studies have suggested peak sensitivity ($\lambda_{\max}$) in the range 455–484 nm^{20–22}. To assess the correspondence, if any, between the Neuro-2a light response and these action spectra, we measured cellular responses to stimuli ranging from 300 to 540 nm in cells transfected with a pCI-neo expression vector driving human melanopsin. EGFP was not included as a marker of expression for these experiments because of its photic properties. Instead, successful transfection was determined by a functional cellular response to the test wavelength or, where that proved ineffective, to a subsequent 420-nm stimulus.

Cells that failed to show a robust physiological response to both stimuli were excluded from subsequent analysis. Cells were preincubated with 11-*cis*-retinaldehyde, although similar results were obtained using 9-*cis*-retinaldehyde as a chromophore (data not shown).

Stimuli at 360 and 420 nm were similarly effective at inducing a cellular response (Fig. 3). Responses at shorter wavelengths were greatly reduced (data not shown), although the significance of this finding is uncertain because the available light at these wavelengths was also substantially attenuated. In contrast, responses to longer wavelengths were impaired despite being associated with increases in light intensity (Fig. 3a), indicating a short wavelength shift in comparison with the human action spectra. In view of the potential for differing melanopsin expression levels between cells under transient transfection, this result was confirmed by direct comparison of responses to 420-nm and 480-nm light in the same cell ($n = 7$). The shorter wavelength was clearly more effective in all of the cells tested (Fig. 3b, c), irrespective of the order of stimulus presentation ($n = 2$ using 420 nm first, $n = 5$ using 480 nm first). We conclude that under these conditions, human melanopsin has a $\lambda_{\max}$ in the range 360–430 nm.

Previous attempts to describe the spectral sensitivity of melanopsin, using purified preparations of mouse melanopsin harvested from COS cells, have reported light-absorbing complexes with $\lambda_{\max}$ around 420 or 440 nm, depending on light exposure history²³.

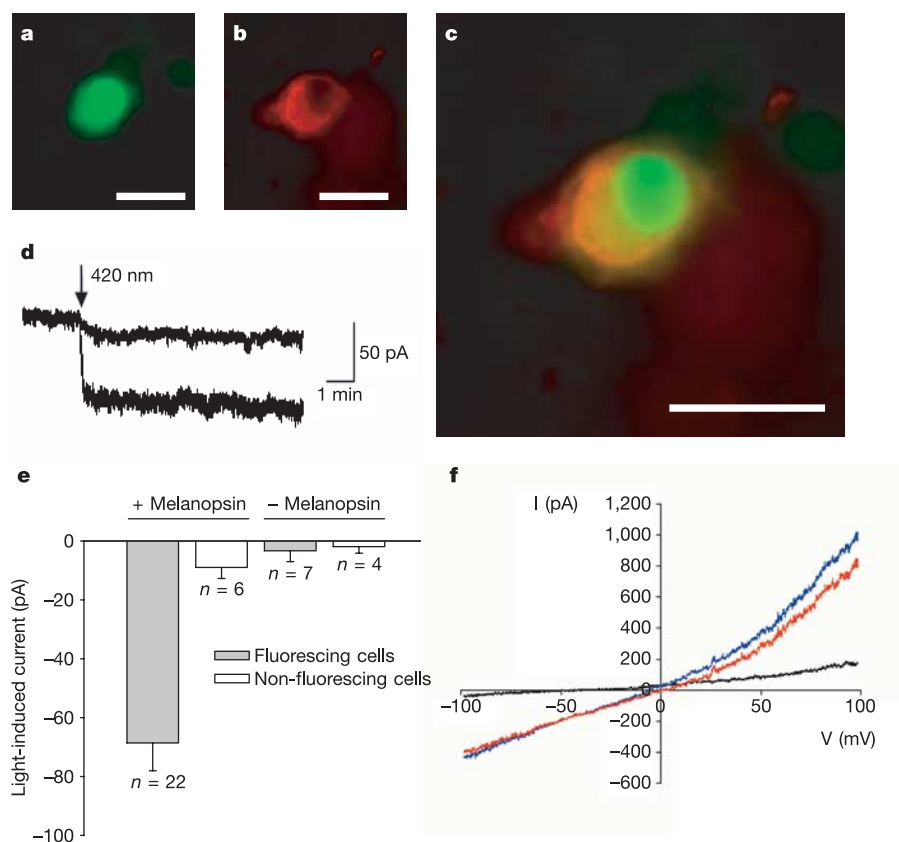


Figure 1 Expression of human melanopsin renders Neuro-2a cells light-sensitive. **a–c**, Neuro-2a cell transfected with 6 × His-tagged human melanopsin-EGFP vector shows EGFP fluorescence (**a**) and anti-tetra-His immunoreactivity associated with His-tagged melanopsin protein (**b**). **c**, Seen in composite; scale bars, 20 μm. **d**, Whole cell recordings from a melanopsin transfected cell preincubated with 9-*cis*-retinaldehyde exposed to 420-nm light (10 s) at 8×10^{13} or 8×10^{14} photons cm⁻² s⁻¹ revealed an intensity-dependent inward current. **e**, Light-evoked currents (mean ± s.e.m.) were

observed in cells successfully transfected with melanopsin-EGFP (+melanopsin, fluorescing) but not in cells transfected with the EGFP vector alone (–melanopsin). **f**, The light-activated current (red) in a melanopsin-expressing cell, determined by subtracting currents elicited by a command voltage ramp (–100 to +100 mV, 2 s) before (black) and during (blue) illumination was inward and linear (–100 to –50 mV), with slope conductance of 6.46 ± 0.82 nS (mean ± s.e.m., $n = 13$).

The data shown here indicate that the physiologically relevant isomerization event for human melanopsin has a $\lambda_{\max}$ around 420 nm or shorter. The red-shifted photoreversal event described above (Fig. 2) may shift this $\lambda_{\max}$ to slightly longer wavelengths when using stimuli of longer duration (>10 min) such as in the human action spectra studies, but it is unlikely to account for the full discrepancy. Of the remaining explanations, parsimony favours the hypothesis that some aspect of the intracellular environment of ipRGCs shifts the spectral sensitivity of melanopsin to longer wavelengths or screens light of shorter wavelength. This might be brought about by a subtle conformational change in the protein, but the functional light response in Neuro-2a cells suggests that the structure of melanopsin in this environment is substantially appropriate.

Photosensory opsins in both vertebrate and invertebrate photo-

receptors are G-protein-coupled. The mammalian melanopsins retain conserved structural features of G-protein-coupled receptors²⁴ and can interact with rod transducin *in vitro*²³. However, putative G-protein interaction domains are highly divergent among melanopsins from different vertebrate species²⁵. The importance of G-protein signalling for human melanopsin function was demonstrated in these experiments by the ability of suramin²⁶ (100 μ M in the bath) to abolish light responses, and confirmed by the observation of an inward current following GTP γ S²⁷ administration (1 mM in patch pipette) that precluded subsequent light responses (Fig. 4a). However, human melanopsin did not appear to be coupled to the classical mammalian photoreceptor G-proteins in Neuro-2a cells, as neither rod nor cone transducins were expressed

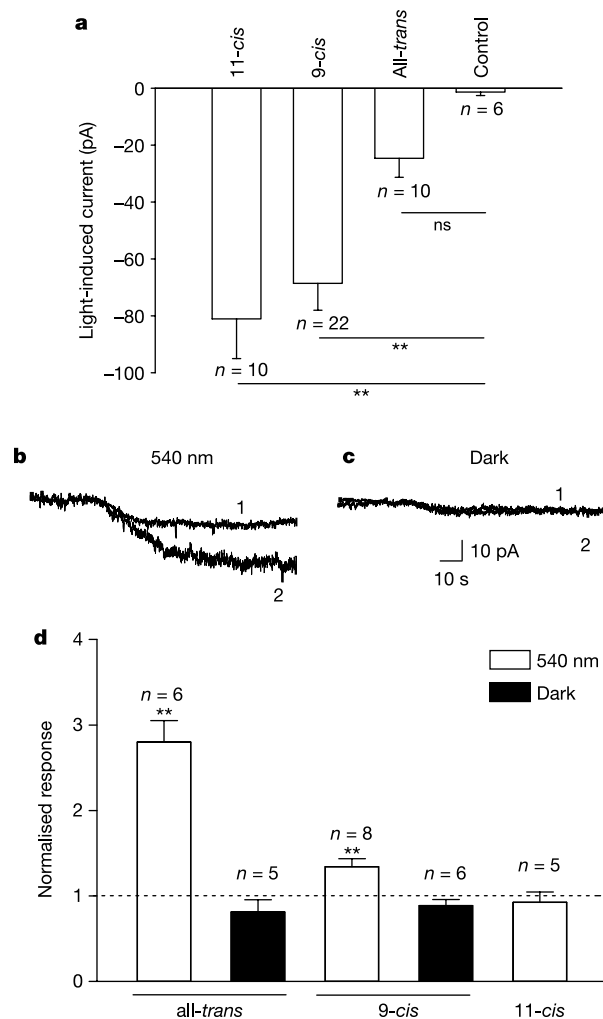


Figure 2 Melanopsin uses *cis*-isoforms of retinaldehyde. **a**, Light-evoked currents (420 nm, 8×10^{14} photons $\text{cm}^{-2} \text{s}^{-1}$ stimulus) were greater in melanopsin-expressing (fluorescent) cells preloaded with 11-*cis*- or 9-*cis*- ($P < 0.001$) but not all-*trans*-retinaldehyde compared with no retinaldehyde controls (mean $\pm$ s.e.m.; ANOVA $P < 0.0001$, post hoc Dunnett's test). The evoked current measured in an all-*trans*-retinaldehyde-loaded cell (**b**, **c**; trace 1) was enhanced following 10 min recovery (in the dark) and a further 10 min exposure to 540-nm light (2×10^{14} photons $\text{cm}^{-2} \text{s}^{-1}$) (**b**; trace 2), but not by 20 min in the dark (**c**; trace 2). **d**, 540-nm treatment caused an increased light response (fold change over baseline response; mean $\pm$ s.e.m.) in cells loaded with all-*trans*- or 9-*cis*- ($P < 0.01$; paired *t*-test) but not 11-*cis*-retinaldehyde, and not observed in cells held in darkness.

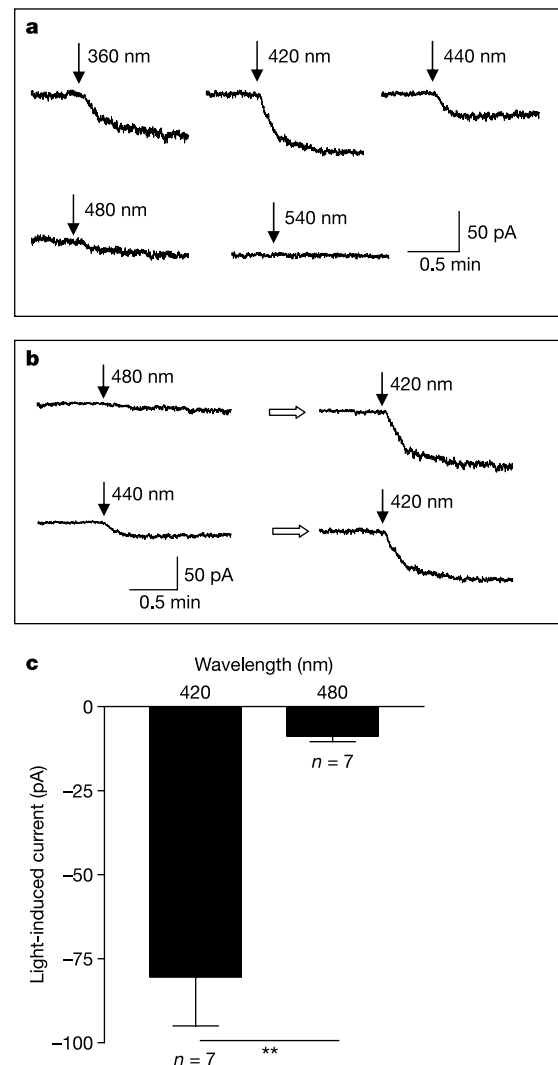


Figure 3 Spectral sensitivity. **a**, Representative responses of cells expressing $6 \times$ His-tagged melanopsin, preloaded with 11-*cis*-retinaldehyde, to 360-nm, 420-nm, 440-nm, 480-nm and 540-nm stimuli (at 3×10^{14} , 8×10^{14} , 2×10^{15} , 3×10^{15} and 10^{14} photons $\text{cm}^{-2} \text{s}^{-1}$ respectively). The 360-nm and 420-nm stimuli gave near-equivalent responses (mean $\pm$ s.e.m., 72.5 ± 19 and 80 ± 11 pA, $n = 6$ and 12 , respectively) whereas responses to the longer wavelengths were progressively attenuated: 32.7 ± 14 pA at 440 nm, 8.5 ± 1.2 pA at 480 nm and 2.0 ± 1.2 pA at 540 nm ($n = 6$, 12 and 4). **b**, Consecutive responses from single cells to 420-nm and 440-nm or 480-nm stimuli confirmed that the shorter wavelength was more effective. **c**, A direct comparison between paired responses to 480-nm and 420-nm stimuli in 7 different cells reinforced this conclusion (paired *t*-test, $P < 0.01$).

in these cells (RT-PCR, data not shown), and pharmacological block of G_i/G_o pathways (using NF023²⁶ or N-ethylmaleimide²⁸) did not antagonize the light response (Fig. 4a).

As human melanopsin acts as a G-protein-coupled receptor, the details of its intracellular transduction cascade are likely to be host-cell-specific. In Neuro-2a cells, we found evidence for the involvement of both intracellular calcium mobilization and cyclic GMP (Fig. 4b, c). Thus, thapsigargin (5–10 μ M, applied for 15–20 min in bath) and 8-Br-cGMP (1 mM in pipette) abolished the light

response. Specific antagonists of phospholipase C or protein kinases A and C did not block the light response (Fig. 4c). The nature of the ion channel regulated by melanopsin activity was investigated by replacement of sodium in the perfusate, removal of calcium and blockade of voltage gated calcium channels, none of which significantly altered the light response (Fig. 4 and data not shown).

These experiments have shown for the first time that expression of melanopsin is not only necessary¹⁴ but also sufficient to render cells functionally photoreceptive. They confirm the ability of human melanopsin to act as a functional sensory photopigment that uses *cis*-isoforms of retinaldehyde and couples to a G-protein signalling pathway. Importantly, the melanopsin photopigment also appears to have an intrinsic mechanism for the regeneration of bleached chromophore. These findings suggest a simple model for the photobiology of ipRGCs, in which a single protein (melanopsin) subserves both sensory and regeneration functions. Under these circumstances, light would drive both photopigment bleach and bleach-recovery processes in a direct manner that has not previously been observed for a mammalian photoreceptor but is common in invertebrates¹⁶. This could explain one of the more surprising observations regarding ipRGCs: their apparent resistance to bleaching under sustained illumination^{1,4}.

An important aspect of this work is the demonstration that, in at least some mammalian cells, a functional photoreceptor can be created by the introduction of only two components: human melanopsin and retinaldehyde. This raises the prospect of using melanopsin expression to render cells photosensitive in a wide variety of experimental and, perhaps, therapeutic applications²⁹.

Note added in proof: While in proof, our attention was drawn to two other papers reporting complementary findings regarding melanopsin function^{31,32}. □

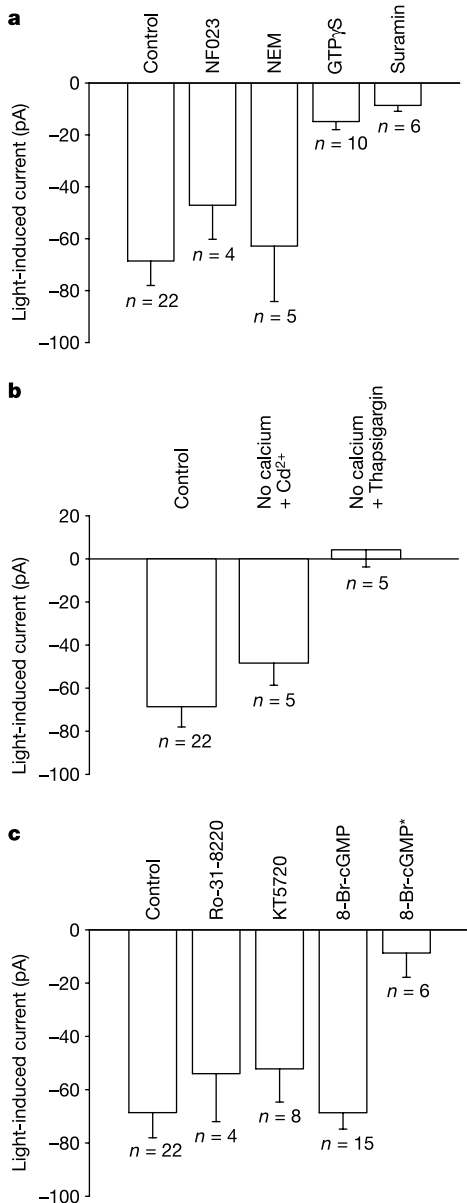


Figure 4 The phototransduction cascade in melanopsin-expressing cells loaded with 9-*cis*-retinaldehyde. **a**, Interruption of G-protein signalling with either GTP γ S (1 mM, in patch pipette) or suramin (100 μ M, in bath) impaired light (420 nm, 10 s, 8×10^{14} photons $\text{cm}^{-2} \text{s}^{-1}$) responses (mean $\pm$ s.e.m.), but neither NF023 nor NEM (1 μ M, 50 μ M in bath, respectively) had an effect. **b**, Removal of calcium from the perfusate and/or the addition of Cd²⁺ (100 μ M) had little effect, while thapsigargin (5–10 μ M for 15–20 min) in zero calcium perfusate abolished responses. Data are means $\pm$ s.e.m. **c**, Neither RO31-8220 (100 nM) nor KT5720 (1 μ M) impaired the light-induced current (mean $\pm$ s.e.m.), but there was a dose-dependent effect of 8-Br-cGMP (in patch pipette, 8-Br-cGMP = 100 μ M; 8-Br-cGMP* = 1 mM).

Methods

Generation of constructs

The full length coding sequence of human melanopsin was amplified from retinal complementary DNA using Pfu-Turbo DNA polymerase (Stratagene) and the following restriction-tagged primers: HMELF, 5'-CCGGAATTCATCCCACTCAGGATGAACC-3' and HMELR, 5'-TGCTCTAGACGCTCTACATCTGGGGTC-3'. The product was sequenced to confirm that there was no deviation from the reported sequence (RefSeq accession number NM_033282) and cloned into the pCMS-EGFP vector (BD Biosciences) so that expression would be driven by the CMV immediate early promoter. A C-terminal, 6 $\times$ His-tagged melanopsin construct (used for the data in Figs 1b, c and 3) was generated in a similar manner, using the following primers: OPN4/5', 5'-GAAGATCTCATCCAA CTCAGGATGAACCCTC-3' and OPN4/3' PH, 5'-CGGAATTCCTAGTGATGGTGAT GGTGATGCATCCTGGGGTCTGGCTGGGGATCAG-3'. We found no discernable difference in the magnitude, duration or spectral sensitivity of responses associated with tagged and untagged melanopsin. For analyses of spectral sensitivity, EGFP expression was excluded by using the pCI-neo mammalian expression vector (Promega).

Cell culture and intracellular recordings

Mouse Neuro-2a cells (CCL-131, American Type Culture Collection) were maintained at 37 $^{\circ}$ C in DMEM medium with GlutaMAX I, 4500 mg l^{-1} D-glucose, sodium pyruvate (DMEM, Invitrogen) and 10% fetal bovine serum, 1% non-essential amino acids and 20 $\mu\text{g ml}^{-1}$ gentomycin in a 5% CO₂ atmosphere. Cells were trypsinized, split and cultured at 1×10^5 cells per ml in 90-mm dishes for 24 h before transfection with plasmid vectors using lipofectamine (Invitrogen) in serum-free medium. The DNA:lipid ratio was 10 μg DNA:100 μg lipid. After transfection, serum-free medium was replaced with normal medium. The next day (2–3 days before study), cell differentiation was induced using 10 μM retinoic acid¹⁶. Cells were kept in darkness starting from this point onwards. On the day of recording, glass coverslips (serving as a substrate for cell attachment) were transferred to the recording chamber. The perfusion solution contained 140 mM NaCl, 4 mM KCl, 1 mM MgCl₂, 2 mM CaCl₂, 5 mM glucose, 10 mM HEPES (pH 7.3–7.4; 24 $^{\circ}$ C). At this point, successfully transfected cells were identified by EGFP fluorescence under a 480-nm stimulus light using an OLYMPUS BX51WI microscope. Retinaldehyde isoforms (20 μM ; 9-*cis*-Retinal and all-*trans*-Retinal from Sigma-Aldrich; 11-*cis*-retinaldehyde from R. K. Crouch) were then added to the perfusion solution as necessary and cells were kept in the dark for at least 1 h before recording. Whole-cell patch-clamp recordings were made with pipettes containing 140 mM KCl, 10 mM NaCl, 1 mM MgCl₂, 10 mM HEPES, 10 mM EGTA. Osmolarity was adjusted to 285 ± 5 mosM and pH to 7.3–7.4 using KOH. Open pipette resistance was 2–5 M Ω , and access resistance during recordings was <20 M Ω . Currents were recorded (Axopatch 200B, Axon Instruments) in neurons that were voltage-clamped at holding potentials of -50 mV. The records were filtered at 1 kHz and sampled at 20 kHz. Drugs were obtained from Sigma Aldrich, with the exception of

N-ethylmaleimide (NEM, Calbiochem), and DL-AP5, NF023, KT5720 and thapsigargin (all Tocris Cookson). Drugs used in superfusion experiments were applied for 15–20 min before light stimulation.

Light stimuli were generated using a Cairn Optoscan Xenon arc source comprising a slit monochromator. Unless otherwise stated, all stimuli were 10 s in duration with a 20-nm half-bandwidth. Irradiance was measured using an optical powermeter (Macam Photometrics) and converted to photon flux. In view of the sustained nature of light responses, analysis was restricted to a single light response from each cell for all of the data except that depicted in Fig. 2b–d and 3b, c, in which paired responses for each cell were recorded. Although subsequent stimuli were presented before full recovery in these instances, we did not find a significant effect on the response (Fig. 2d). The magnitude of the responses was defined by the peak sustained current, measured using Clampfit (Axon Instruments). It has been suggested that NMDA receptors may include a light sensitive moiety³⁰. The nature of glutamate receptor expression in Neuro-2a cells remains ambiguous, nonetheless we excluded the possibility that this might be the origin of the melanopsin light response by application of the selective NMDA antagonist DL-AP5 (100 μ M), which had no effect on the light sensitive current (data not shown).

Opsin expression in Neuro-2a cells by RT-PCR

Cells were collected both before and after differentiation and after transfection with pCMS-EGFP vector alone or the human melanopsin-EGFP vector. RNA was extracted using Tri reagent (Sigma) and treated with DNaseI (Promega) before reverse transcription. Single-stranded cDNA was synthesized using a SuperScript kit (Invitrogen). Specific primers that amplify both mouse and human melanopsin were used, as well as species-specific primers for mouse rod opsin, UVS/MWS-cone opsins, RGR-opsin, peropsin and rod/cone transducin α subunits (Gnat1 and Gnat2). The absence of vector DNA carry-over in both of the transfected Neuro-2a cDNA samples was confirmed using primers designed to the CMV promoter (data not shown).

Immunocytochemistry

Cells were fixed with 4% paraformaldehyde in PBS buffer and exposed to a primary Tetra-His antibody (Qiagen) at 1:100 dilution. Visualization was achieved by fluorescence microscopy (Nikon Eclipse E600; MetaMorph, Universal Imaging Corp) after incubation with a tetramethylrhodamine goat anti-mouse secondary antibody at 1:400.

Received 21 October 2004; accepted 12 January 2005; doi:10.1038/nature03344.
Published online 26 January 2005.

- Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
- Hattar, S. *et al.* Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. *Nature* **424**, 75–81 (2003).
- Panda, S. *et al.* Melanopsin is required for non-image-forming photic responses in blind mice. *Science* **301**, 525–527 (2003).
- Sekaran, S., Foster, R. G., Lucas, R. J. & Hankins, M. W. Calcium imaging reveals a network of intrinsically light-sensitive inner-retinal neurons. *Curr. Biol.* **13**, 1290–1298 (2003).
- Warren, E. J., Allen, C. N., Brown, R. L. & Robinson, D. W. Intrinsic light responses of retinal ganglion cells projecting to the circadian system. *Eur. J. Neurosci.* **17**, 1727–1735 (2003).
- Lucas, R., Douglas, R. & Foster, R. Characterization of an ocular photopigment capable of driving pupillary constriction in mice. *Nature Neurosci.* **4**, 621–626 (2001).
- Lucas, R. J. *et al.* Identifying the photoreceptive inputs to the mammalian circadian system using transgenic and retinally degenerate mice. *Behav. Brain Res.* **125**, 97–102 (2001).
- Czeisler, C. A. *et al.* Suppression of melatonin secretion in some blind patients by exposure to bright light. *N. Engl. J. Med.* **332**, 6–11 (1995).
- Mrosovsky, N., Lucas, R. & Foster, R. Persistence of masking responses to light in mice lacking rods and cones. *J. Biol. Rhythms* **16**, 585–587 (2001).
- Lockley, S. W. *et al.* Relationship between melatonin rhythms and visual loss in the blind. *J. Clin. Endocrinol. Metab.* **82**, 3763–3770 (1997).
- Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Gooley, J. J., Lu, J., Chou, T. C., Scammell, T. E. & Saper, C. B. Melanopsin in cells of origin of the retinohypothalamic tract. *Nature Neurosci.* **4**, 1165 (2001).
- Hannibal, J., Hindersson, P., Knudsen, S. M., Georg, B. & Fahrenkrug, J. The photopigment melanopsin is exclusively present in pituitary adenylate cyclase-activating polypeptide-containing retinal ganglion cells of the retinohypothalamic tract. *J. Neurosci.* **22**, RC191 (2002).
- Lucas, R. J. *et al.* Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).
- Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P. & Rollag, M. D. Melanopsin: An opsin in melanophores, brain, and eye. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
- Hubbard, R. & St. George, R. The rhodopsin system of the squid. *J. Gen. Physiol.* **41**, 501–528 (1958).
- Liu, R. S., Crescitelli, E., Denny, M., Matsumoto, H. & Asato, A. E. Photosensitivity of 10-substituted visual pigment analogues: detection of a specific secondary opsin-retinal interaction. *Biochemistry* **25**, 7026–7030 (1986).
- Hao, W. & Fong, H. The endogenous chromophore of retinal G protein-coupled receptor opsin from the pigment epithelium. *J. Biol. Chem.* **274**, 6085–6090 (1999).
- Koyanagi, M. *et al.* Bistable UV pigment in the lamprey pineal. *Proc. Natl Acad. Sci. USA* **101**, 6687–6691 (2004).
- Brainard, G. *et al.* Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *J. Neurosci.* **21**, 6405–6412 (2001).
- Thapan, K., Arendt, J. & Skene, D. An action spectrum for melatonin suppression: evidence for a novel non-rod, non-cone photoreceptor system in humans. *J. Physiol. (Lond.)* **535**, 261–267 (2001).
- Hankins, M. W. & Lucas, R. J. The primary visual pathway in humans is regulated according to long-term light exposure through the action of a nonclassical photopigment. *Curr. Biol.* **12**, 191–198 (2002).

- Newman, L. A., Walker, M. T., Brown, R. L., Cronin, T. W. & Robinson, P. R. Melanopsin forms a functional short-wavelength photopigment. *Biochemistry* **42**, 12734–12738 (2003).
- Provencio, I. *et al.* A novel human opsin in the inner retina. *J. Neurosci.* **20**, 600–605 (2000).
- Bellingham, J., Whitmore, D., Philp, A. R., Wells, D. J. & Foster, R. G. Zebrafish melanopsin: isolation, tissue localisation and phylogenetic position. *Brain Res. Mol. Brain Res.* **107**, 128–136 (2002).
- Beindl, W. *et al.* Inhibition of receptor/G protein coupling by suramin analogues. *Mol. Pharmacol.* **50**, 415–423 (1996).
- Gilman, A. G proteins and dual control of adenylate cyclase. *Cell* **36**, 577–579 (1984).
- Asano, T. & Ogasawara, N. Uncoupling of gamma-aminobutyric acid B receptors from GTP-binding proteins by N-ethylmaleimide: effect of N-ethylmaleimide on purified GTP-binding proteins. *Mol. Pharmacol.* **29**, 244–249 (1986).
- Zemelman, B. V., Lee, G. A., Ng, M. & Miesenböck, G. Selective photostimulation of genetically chARGed neurons. *Neuron* **33**, 15–22 (2002).
- Leszkiewicz, D., Kandler, K. & Aizenman, E. Enhancement of NMDA receptor-mediated currents by light in rat neurones *in vitro*. *J. Physiol. (Lond.)* **524**, 365–374 (2000).
- Qiu, X. *et al.* Induction of photosensitivity by heterologous expression of melanopsin. *Nature* doi:10.1038/nature03345 (this issue).
- Panda, S. *et al.* Illumination of the melanopsin signaling pathway. *Science* **307**, 600–604 (2005).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by a Wellcome Trust Showcase Award (M.W.H. and R.J.L.) and in part by the BBSRC (R.J.L.) and NSBRI through NASA NCC 9-58 (R.J.L. and R. Foster). We are grateful to R.K. Crouch for the gift of 11-*cis* retinaldehyde, to R. Douglas for scientific discussions and comments on the manuscript, and to K. Wells for help and advice with cell culture.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.J.L. (robert.lucas@manchester.ac.uk) or M.W.H. (m.hankins@imperial.ac.uk).

Induction of photosensitivity by heterologous expression of melanopsin

Xudong Qiu^{1*}, Tida Kumbalasisri^{2*}, Stephanie M. Carlson¹, Kwoon Y. Wong¹, Vanitha Krishna¹, Ignacio Provencio^{2†} & David M. Berson¹

¹Department of Neuroscience, Box 1953, Brown University, Providence, Rhode Island 02912-1953, USA

²Department of Anatomy, Physiology and Genetics and the Graduate Program in Neuroscience, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814, USA

* These authors contributed equally to this work

† Current address: Department of Biology, University of Virginia, Charlottesville, Virginia 22904, USA

Melanopsin^{1–8} has been proposed to be the photopigment of the intrinsically photosensitive retinal ganglion cells (ipRGCs)^{7–15}; these photoreceptors of the mammalian eye drive circadian and pupillary adjustments through direct projections to the brain^{5,6,8–14,16–18}. Their action spectrum ($\lambda_{\max} \approx 480$ nm) implicates an opsin¹⁰ and melanopsin is the only opsin known to exist in these cells. Melanopsin is required for ipRGC photosensitivity¹³ and for behavioural photoresponses that survive disrupted rod and cone function^{14,17}. Heterologously expressed melanopsin apparently binds retinaldehyde and mediates photic activation of G proteins¹⁹. However, its amino-acid sequence differs from vertebrate photosensory opsins^{1,20} and some have suggested that melanopsin may be a photoisomerase, providing retinoid chromophore to an unidentified opsin^{3,20}. To determine whether melanopsin is a functional sensory photopigment, here we transiently expressed it in HEK293 cells that stably expressed TRPC3 channels. Light triggered a membrane depolarization in these cells and increased intracellular calcium. The light

response resembled that of ipRGCs, with almost identical spectral sensitivity ($\lambda_{\max} \approx 479$ nm). The phototransduction pathway included Gq or a related G protein, phospholipase C and TRPC3 channels. We conclude that mammalian melanopsin is a functional sensory photopigment, that it is the photopigment of ganglion-cell photoreceptors, and that these photoreceptors may use an invertebrate-like phototransduction cascade.

To test melanopsin's capacity to form a functional sensory photopigment, we heterologously expressed mouse *melanopsin* (*Opn4*) complementary DNA in human embryonic kidney cells (HEK293) stably expressing TRPC3 receptor-operated non-specific cation channels²¹. The mouse melanopsin open reading frame was cloned into a bicistronic expression vector (pIRES2-EGFP). HEK293-TRPC3 cells transfected with this vector faithfully co-expressed the enhanced green fluorescent protein (EGFP) reporter protein (Fig. 1a, c) and mouse melanopsin (Fig. 1b, c). Melanopsin was localized predominantly in the cell membrane (Fig. 1b), as it is in ipRGCs²².

Many of these cells were photosensitive. They exhibited large light-evoked changes in transmembrane voltage that were sluggish and sustained (Figs 1d, e). Light responses were never observed among neighbouring untransfected, EGFP-negative cells ($n = 11$),

nor in cells transiently transfected with empty pIRES2-EGFP, which expressed EGFP but not melanopsin ($n = 10$; Fig. 1f). Increasing stimulus irradiance augmented response amplitude (Fig. 1e; see also Fig. 4a) and decreased onset latency from >10 s to a minimum of ~ 500 ms. Maximal responses were typically >15 mV. Thresholds were approximately $(5.6 \pm 1.7) \times 10^{12}$ photons $s^{-1} cm^{-2}$ at 480 nm (mean $\pm$ standard error of the mean, s.e.m.; $n = 12$), about one log unit higher than for ipRGCs¹⁰. Although there was no retinaldehyde in the bath, responsiveness persisted for hours.

With voltage clamped at negative potentials, light triggered an inward current (Fig. 1g) exhibiting strong outward rectification and reversing near 0 mV (4.0 ± 3.8 mV, mean $\pm$ s.e.m., $n = 9$; Fig. 1h). Maximal white-light stimuli evoked peak inward currents averaging 30 pA (± 13 pA standard deviation, s.d.; $n = 5$; $V_{\text{hold}} \approx -44$ mV). Light-evoked inward currents were never observed in untransfected cells (Fig. 1h, bottom) or in cells expressing EGFP but not melanopsin.

Light also triggered increases in intracellular free calcium in these cells. Only transfected cells exhibited light-evoked increases in fluorescence of the calcium indicator Rhod-2-AM (Fig. 2). Calcium responses were also detectable in melanopsin-transfected cells lacking the EGFP reporter (Supplementary Fig. 1).

Melanopsin is presumably the photopigment in these cells

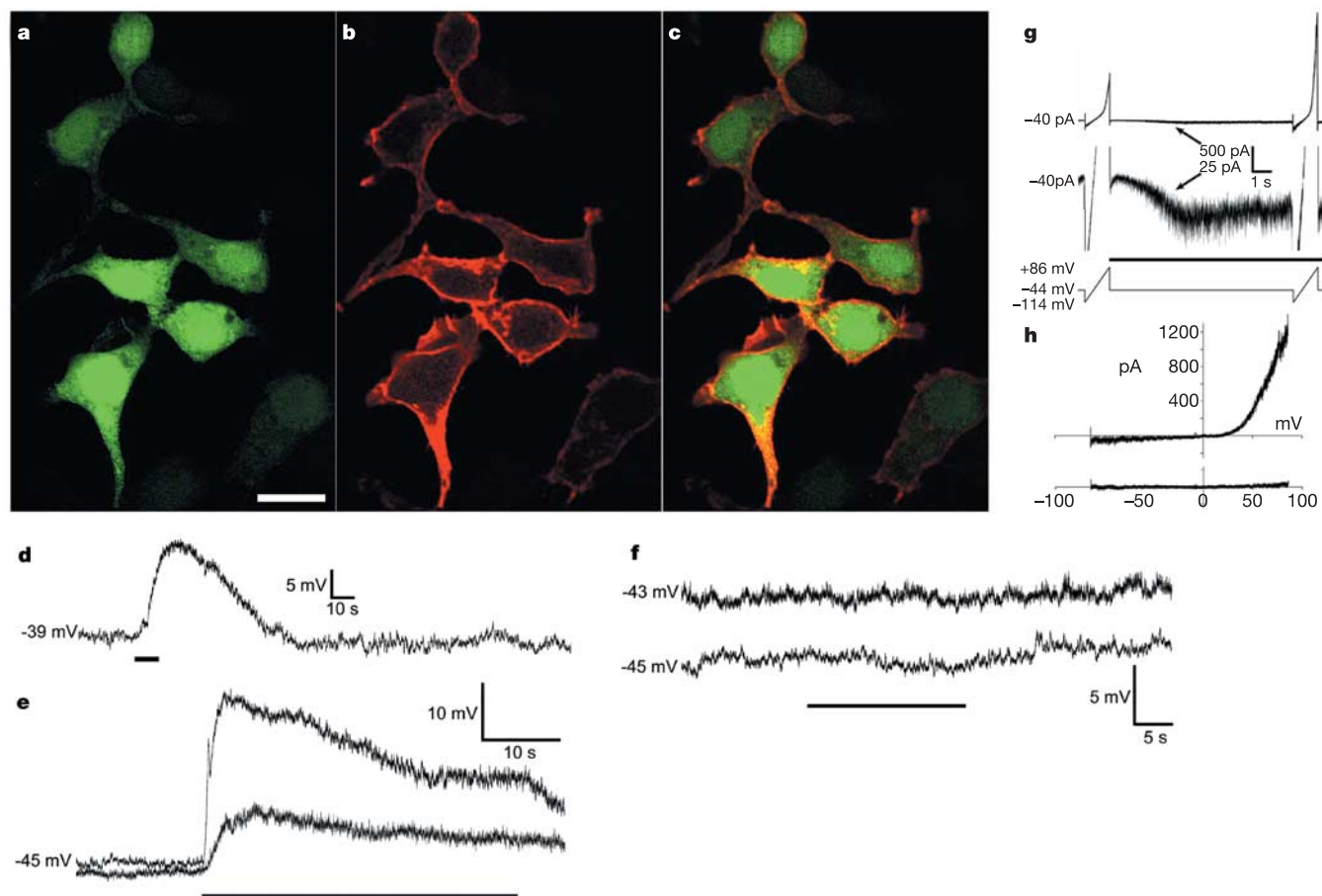


Figure 1 Light-evoked responses of melanopsin-expressing HEK293-TRPC3 cells. **a–c**, Melanopsin expression. Confocal photomicrographs showing EGFP (**a**, green), melanopsin immunofluorescence (**b**, red) and merge (**c**). Scale bar, 20 μ m. **d**, **e**, Current-clamp recordings of light-evoked depolarization in transfected cells. **d**, Sluggish response to 10-s flash (bar marks stimulus; 480 nm, 9×10^{14} photons $s^{-1} cm^{-2}$). **e**, Intensity coding. The white-light stimulus for top trace was one log unit more intense than for bottom trace. **f**, Absence of photosensitivity in untransfected cells (top trace) and in cells

expressing EGFP but not melanopsin (bottom trace). **g**, **h**, Voltage-clamp recordings. **g**, The light stimulus (horizontal bar) elicits tonic inward current and increased current noise (middle trace). The same trace at lower gain (top trace) shows responses to voltage ramps (bottom trace) imposed before and during light response. **h**, Current–voltage relations of light-evoked current in **g** (stimulated minus dark current) (top trace). Absence of light-evoked current in untransfected cell (bottom trace).

because it was required for the observed photosensitivity. To test the hypothesis that melanopsin, like other opsin photopigments, triggers a G-protein signalling cascade, we infused guanosine 5'-O-(2-thiodiphosphate) (GDP β S). Light-evoked depolarizations were reduced by more than half their initial amplitude within 30 min (Fig. 3a; in 3 of 4 cells tested). This was a specific effect, not an artefact of cell dialysis, because cells recorded with the control internal solution maintained stable response amplitudes for ≥ 70 min (data not shown). Gq or a related protein seems most likely to be the cognate G protein for melanopsin. The light response was effectively abolished within 20 min by internal application of GPant-2a, a competitive inhibitor of Gq, and was suppressed 75–100% by bath-application of U73122, an antagonist of phospholipase C (PLC; $n = 3$; Fig. 3c), the effector enzyme for Gq-like G proteins²³. An inactive analogue of the PLC antagonist (U73343; $n = 3$; Supplementary Fig. 2a) and pertussis toxin (a Gi inactivator; 250 ng ml⁻¹, 2 h) were ineffective. TRPC3 channels apparently carry the light-activated current in these cells (see Supplementary Data and Supplementary Figs 2b and c).

Relative sensitivity at different wavelengths was assessed for individual cells as shown in Fig. 4a. The optimal wavelength ($\lambda_{\max}$) for each cell was estimated from the best-fitting retinaldehyde template function²⁴ (see Methods). These values, plotted in the histogram of Fig. 4b, had a mean value of 479.2 nm (± 1.5 nm s.e.m.; $n = 12$), very close to the optimal wavelength for isolated rat ipRGCs¹⁰ (484 nm). Mean estimates of relative sensitivity at each tested wavelength (Fig. 4b, points) adhered closely to the best-fitting retinaldehyde template function (Fig. 4b, curve), supporting the

assumption that the action spectrum in this system adheres to the standard form for opsin-mediated responses.

This study provides the first direct physiological evidence that melanopsin is a functional sensory photopigment and that it can activate a signalling cascade that gates an ionic conductance. Melanopsin is an opsin, with the canonical retinaldehyde binding site^{1,20}, and most cells expressing it are known or presumed to be directly photosensitive^{1,2,6–8,10–14}. Its overexpression increases photosensitivity in dermal melanophores²⁵, and its deletion abolishes the light response in ganglion-cell photoreceptors¹³. Heterologously expressed melanopsin reportedly mediates photic activation of a G protein¹⁹, again supporting a photosensory role.

Our findings contradict the idea that melanopsin could be exclusively a photoisomerase instead of a photosensory pigment^{3,20}. Introducing melanopsin into HEK293-TRPC3 cells transformed them into photoreceptors. The action spectrum of the induced photoresponse peaked at 479 nm, closely matching the spectral tuning of isolated rat ipRGCs¹⁰ ($\lambda_{\max} \approx 484$ nm) and of melanopsin-dependent behavioural light responses of rodless/coneless

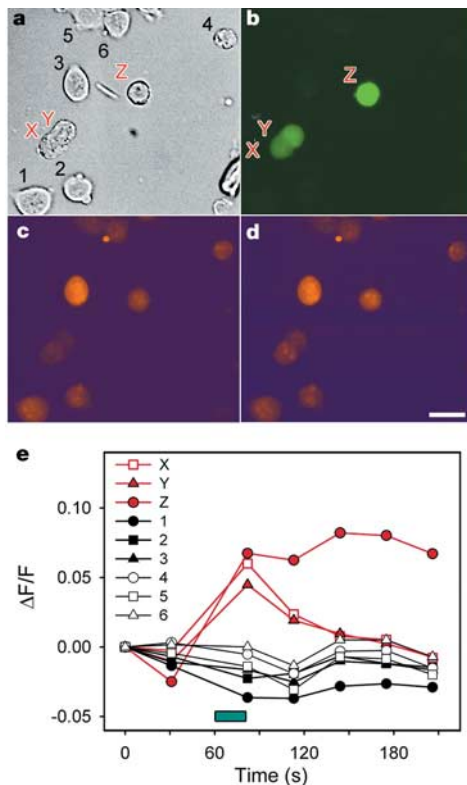


Figure 2 Light evokes calcium responses. Brightfield image (a) and EGFP fluorescence (b) of live cells. Of nine cells, three (X, Y and Z) were transfected (EGFP+). c, d, Pseudocolour images of Rhod-2-AM fluorescence before (c) and after (d) exposure to a light stimulus (500 nm, 10^{15} photons s⁻¹ cm⁻²). Only transfected cells exhibited increased calcium signals. Scale bar, 20 μ m. e, Normalized change in fluorescence over time for transfected cells (red lines) and untransfected cells (black lines). The green bar marks the stimulus.

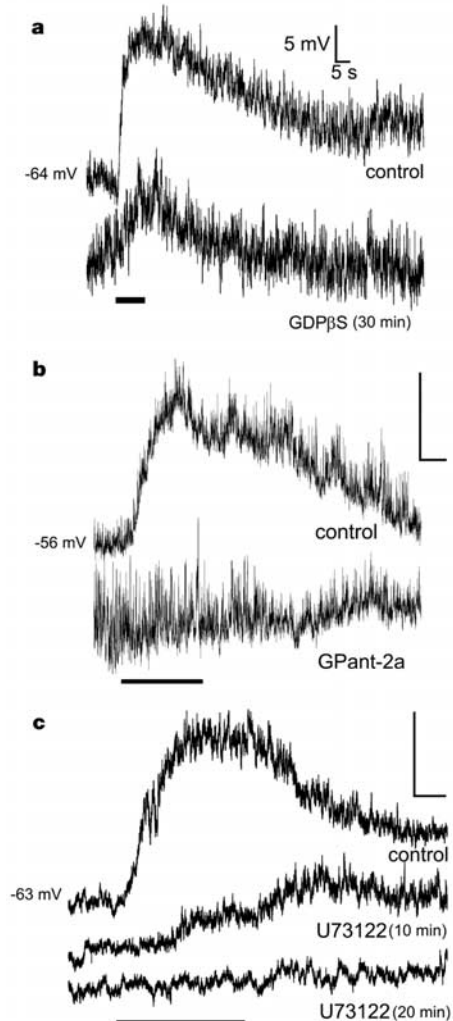


Figure 3 Phototransduction cascade. Voltage responses to bright white-light stimuli (horizontal bars) before and during drug application. Photoresponses were reduced or blocked by internal application of GDP β S (a; 2 mM), a non-specific antagonist of G proteins, or of GPant-2a (40 μ M), a Gq/11 antagonist (b). c, Bath application of the PLC antagonist U73122 (10 μ M) eliminated the light response (c); the inactive analogue U73343 had no effect (Supplementary Fig. 2a). Calibration bars, 5 mV, 5 s.

rodents (pupillary light reflex: $\lambda_{\max} \approx 479$ nm, ref. 16; circadian phase shifting: $\lambda_{\max} \approx 480$ –481 nm, refs 17 and 26). Such correspondence was lacking in an earlier study of heterologously expressed melanopsin, which reported maximal absorption at 421 nm (ref. 20). The discrepancy with our spectral results could stem from the earlier study's use of a different expression system, protein modifications for epitope tagging, effects of opsin solubilization and immunopurification, or the use of chemical rather than photic bleaching to generate the difference spectrum.

Though melanopsin is now firmly established as a photosensory pigment, it may also possess photoisomerase activity. In bistable invertebrate photopigments, the chromophore–opsin linkage is non-dissociating and light can trigger photoreversal of metarhodopsin to rhodopsin. Support for possible melanopsin bistability comes from its invertebrate-like aromatic amino acid at the 'counter-ion' position¹, a key determinant of chromophore binding, and, as shown here, the persistence for hours of melanopsin-based light responses without supplementary retinaldehyde despite repeated exposure to bright illumination. Still, there is no direct evidence that melanopsin is bistable. In photoreceptors using bistable pigments, appropriate narrow-band stimuli can terminate persistent post-stimulus responses and augment sensitivity to subsequent stimuli by converting metarhodopsin to rhodopsin. We have not detected such behaviour in transfected HEK293-TRPC3 cells (see Sup-

plementary Data). If melanopsin instead dissociates from its chromophore like other vertebrate photopigments, its regeneration in HEK293 cells may be attributable to the intrinsic retinoid processing capacity of these cells²⁷.

By demonstrating that melanopsin is activated maximally near 480 nm and can indirectly gate cation channels in the plasma membrane, our findings provide nearly all of the missing links in the chain of evidence for melanopsin as the photopigment of ipRGCs^{8,10,20}. There are striking similarities between the photoresponses of melanopsin-expressing HEK293 cells and those of ipRGCs including response polarity, low sensitivity, sluggish onset, tonic response to continuous illumination, and slow post-stimulus recovery. It seems probable that some of these properties, like the spectral tuning, derive from features of melanopsin itself, while others are shaped by the downstream signalling cascade.

In HEK293-TRPC3 cells, this cascade involves Gq or a related pertussis toxin-insensitive G protein. Some G-protein-coupled receptors (GPCRs) couple promiscuously to multiple G-protein families²⁸ and melanopsin itself reportedly interacts with highly concentrated rod transducin (Gt) in a biochemical assay¹⁹. Under physiological conditions, however, most GPCRs couple mainly to only one of the four G-protein families²⁸, so the cognate G protein for melanopsin in ganglion cells is probably a member of the Gq family²³ (Gq, G11, G14, G15 and/or G16). In HEK293-TRPC3 cells, the G protein activated by melanopsin stimulates PLC to open TRPC3 channels and depolarize the plasma membrane^{21,23,29}. *Drosophila* photoreceptors apparently use a similar phototransduction cascade³⁰, but it is not yet known whether ipRGCs use (or even express) these specific downstream components.

Note added in proof: While in proof we became aware of three other papers to be published that concurrently support our conclusion that melanopsin is a functional photopigment, signalling through G proteins^{31–33}. □

Methods

Cell culture and melanopsin expression

Cells were cultured conventionally. The mouse melanopsin (*Opn4*) open reading frame (GenBank accession number NM_013887) was amplified by polymerase chain reaction (PCR) from a vector containing the open reading frame. PCR primers were generated from the melanopsin sequence with restriction enzyme sequences appended to the 5' end to facilitate directional cloning. Two different PCR products were generated for two expression vectors: the plasmid pcDNA 3.1(+) (Invitrogen) and the bicistronic pIRES2-EGFP (BD Biosciences). PCR products were digested with restriction enzymes and ligated into the multiple cloning site of the vectors, also digested with restriction enzymes. Expression vector sequences were confirmed by dideoxy terminator sequencing. Cells were transfected with the calcium phosphate method and seeded on coverslips. From transfection until recording, cells were provided with retinaldehyde (0.5 μ M 11-*cis*-retinaldehyde for spectral studies, 2 μ M all-*trans*-retinaldehyde for all other experiments) and were kept in darkness, except for brief exposures to laboratory lighting. Omitting these retinoids did not preclude photosensitivity (see Supplementary Data).

Melanopsin immunohistochemistry

Cells were fixed overnight (4% paraformaldehyde in phosphate-buffered saline, PBS), incubated in polyclonal rabbit anti-melanopsin antiserum (UF006, ref. 7, 1:2,500; 24 h, 4 °C), then in Cy-3-conjugated anti-rabbit IgG (1:500; Jackson ImmunoResearch; 1 h; 21 °C). Photomicrographs were obtained on a Zeiss Pascal confocal microscope using a 40 $\times$ objective.

Electrophysiology

Cells transfected 1–2 days earlier were mounted in a chamber perfused with Tyrodes solution containing no retinaldehyde (1–2 ml min⁻¹, 21 °C). Although incubated in darkness, cells were exposed to normal laboratory lighting during preparation for recording except in spectral studies, which were conducted in darkness. Cells were visualized with a 40 $\times$ water-immersion objective on an upright epifluorescence microscope equipped with a cooled charge-coupled device (CCD) camera, integrating frame grabbers and video monitor. Except for brief blue epifluorescence to identify EGFP-positive cells (Chroma no. 41001), only infrared light was used to view cells.

Whole-cell patch recordings were made conventionally. Micropipettes (3–5 M Ω) contained 120 mM K-gluconate, 5 mM NaCl, 4 mM KCl, 0.5–2 mM EGTA, 10 mM HEPES, 0.5–4 mM ATP-Mg, 7 mM phosphocreatine, 0.05–0.3 mM GTP-Tris; (280–300 mOsm, pH 7.3). Liquid junction potentials (–14 mV) were corrected. In darkness, most cells had series resistance <10 M Ω , $V_m \approx -35$ to -45 mV and input resistance of 0.5–1.5 G Ω . Evoked currents or voltages are baseline-subtracted.

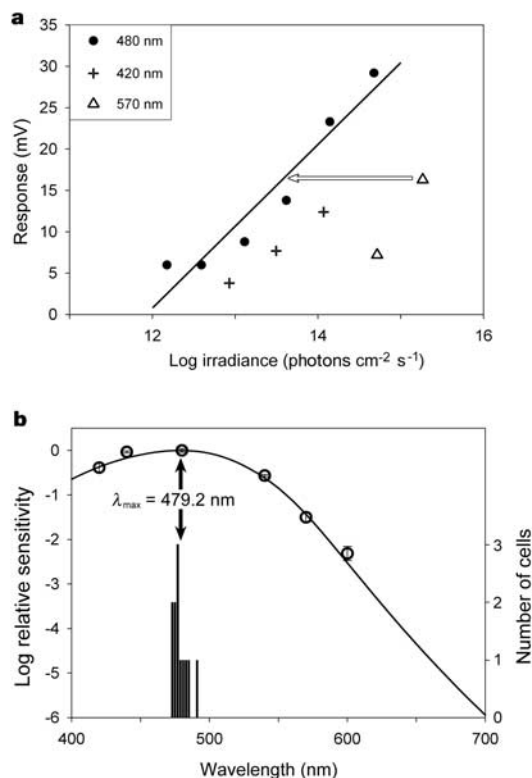


Figure 4 Spectral sensitivity. **a**, Plot of peak depolarization evoked in one cell by stimuli of various wavelengths. The solid line shows the least-squares linear fit to 480-nm data (solid circles). The horizontal displacement of other points from this line (open arrow) estimates sensitivity at corresponding wavelength relative to 480 nm. **b**, The histogram shows the distribution of the estimated optimal wavelength ($\lambda_{\max}$), determined for each cell from the best-fitting retinaldehyde template. Mean $\lambda_{\max} = 479.2$ nm $\pm$ 1.5 nm ($n = 12$). Above the histogram, each point represents the mean value of relative sensitivity at that wavelength among all cells (error bars, barely visible, are s.e.m.). The curve is the least-squares best-fitting retinaldehyde template function; $\lambda_{\max}$ (479.2 nm) has an ordinate value of 0. Medium supplemented with 11-*cis*-retinaldehyde.

Photic stimulation and spectral analysis

Light stimuli from a 100-W tungsten source were filtered (neutral density and narrow-band interference filters; 10 nm width; Oriol), gated by a shutter (Uniblitz VS35; Vincent Associates) and calibrated by a radiometer (S370, UDT Instruments). The irradiance of the unfiltered ('white') stimulus was (in photons $\text{s}^{-1} \text{cm}^{-2}$): 4×10^{12} at 400 nm, 6×10^{13} at 500 nm and 1×10^{14} at 600 nm.

For spectral analysis, the culture medium contained 11-*cis*-retinaldehyde. Current injection held cells near -44 mV. Stable sensitivity, confirmed by retesting with a standard stimulus, required long interstimulus intervals (7 min). Stimuli of 480 nm were interleaved with those of two to three other wavelengths (420, 440, 540, 570 or 600 nm). Each response to one of these other wavelengths yielded an estimate of relative sensitivity normalized to that at 480 nm (see Fig. 4a), and these were averaged for each wavelength. For each cell, we determined the retinaldehyde template function²⁴ that best fitted these data (least-squares method). The two free parameters for the fit were λ_{max} and the vertical offset. Relative sensitivities were then re-normalized to that at the theoretical optimum (λ_{max}).

Calcium imaging

Cells loaded with the long-wavelength Ca^{2+} indicator Rhod-2-AM (4–6 μM in Tyrodes, 30 min, 37 °C) were imaged at 30 Hz using green excitation (525–550 nm; emission 580–650 nm; Chroma 31002a) attenuated 128- to 512-fold by neutral density filters. Integration time was fixed within trials (typically 32 frames) to optimize sensitivity and avoid saturation. Responses, analysed by ImageJ software (W. Rasband; <http://rsb.info.nih.gov/ij/>; 2004), were expressed as post-stimulus change in fluorescence intensity above baseline divided by the baseline intensity ($\Delta F/F$).

Further methodological details are provided in the Supplementary Methods.

Received 5 November 2004; accepted 10 January 2005; doi:10.1038/nature03345.

Published online 26 January 2005.

1. Provencio, I., Jiang, G., De Grip, W. J., Hayes, W. P. & Rollag, M. D. Melanopsin: An opsin in melanophores, brain, and eye. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
2. Provencio, I. *et al.* A novel human opsin in the inner retina. *J. Neurosci.* **20**, 600–605 (2000).
3. Bellingham, J., Whitmore, D., Philp, A. R., Wells, D. J. & Foster, R. G. Zebrafish melanopsin: isolation, tissue localisation and phylogenetic position. *Brain Res. Mol. Brain Res.* **107**, 128–136 (2002).
4. Hannibal, J. & Fahrenkrug, J. Melanopsin: a novel photopigment involved in the photoentrainment of the brain's biological clock? *Ann. Med.* **34**, 401–407 (2002).
5. Hannibal, J., Hindersson, P., Knudsen, S. M., Georg, B. & Fahrenkrug, J. The photopigment melanopsin is exclusively present in pituitary adenylate cyclase-activating polypeptide-containing retinal ganglion cells of the retinohypothalamic tract. *J. Neurosci.* **22**, RC191 (2002).
6. Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
7. Provencio, I., Rollag, M. D. & Castrucci, A. M. Photoreceptive net in the mammalian retina. *Nature* **415**, 493 (2002).
8. Berson, D. M. Strange vision: ganglion cells as circadian photoreceptors. *Trends Neurosci.* **26**, 314–320 (2003).
9. Gooley, J. J., Lu, J., Chou, T. C., Scammell, T. E. & Saper, C. B. Melanopsin in cells of origin of the retinohypothalamic tract. *Nature Neurosci.* **4**, 1165 (2001).
10. Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
11. Panda, S. *et al.* Melanopsin (Opn4) requirement for normal light-induced circadian phase shifting. *Science* **298**, 2213–2216 (2002).
12. Ruby, N. F. *et al.* Role of melanopsin in circadian responses to light. *Science* **298**, 2211–2213 (2002).
13. Lucas, R. J. *et al.* Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).
14. Panda, S. *et al.* Melanopsin is required for non-image-forming photic responses in blind mice. *Science* **301**, 525–527 (2003).
15. Warren, E. J., Allen, C. N., Brown, R. L. & Robinson, D. W. Intrinsic light responses of retinal ganglion cells projecting to the circadian system. *Eur. J. Neurosci.* **17**, 1727–1735 (2003).
16. Lucas, R. J., Douglas, R. H. & Foster, R. G. Characterization of an ocular photopigment capable of driving pupillary constriction in mice. *Nature Neurosci.* **4**, 621–626 (2001).
17. Hattar, S. *et al.* Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. *Nature* **424**, 76–81 (2003).
18. Gooley, J. J., Lu, J., Fischer, D. & Saper, C. B. A broad role for melanopsin in nonvisual photoreception. *J. Neurosci.* **23**, 7093–7106 (2003).
19. Newman, L. A., Walker, M. T., Brown, R. L., Cronin, T. W. & Robinson, P. R. Melanopsin forms a functional short-wavelength photopigment. *Biochemistry* **42**, 12734–12738 (2003).
20. Bellingham, J. & Foster, R. G. Opsins and mammalian photoentrainment. *Cell Tissue Res.* **309**, 57–71 (2002).
21. Hurst, R. S., Zhu, X., Boulay, G., Birnbaumer, L. & Stefani, E. Ionic currents underlying HTRP3 mediated agonist-dependent Ca^{2+} influx in stably transfected HEK293 cells. *FEBS Lett.* **422**, 333–338 (1998).
22. Belenky, M. A., Smeraski, C. A., Provencio, I., Sollars, P. J. & Pickard, G. E. Melanopsin retinal ganglion cells receive bipolar and amacrine cell synapses. *J. Comp. Neurol.* **460**, 380–393 (2003).
23. Fields, T. A. & Casey, P. J. Signalling functions and biochemical properties of pertussis toxin-resistant G-proteins. *Biochem. J.* **321**, 561–571 (1997).
24. Lamb, T. D. Photoreceptor spectral sensitivities: common shape in the long- wavelength region. *Vision Res.* **35**, 3083–3091 (1995).
25. Rollag, M. D., Provencio, I., Sugden, D. & Green, C. B. Cultured amphibian melanophores: a model system to study melanopsin photobiology. *Methods Enzymol.* **316**, 291–309 (2000).
26. Yoshimura, T. & Ebihara, S. Spectral sensitivity of photoreceptors mediating phase-shifts of circadian rhythms in retinally degenerate CBA/J (rd/rd) and normal CBA/N (+/+) mice. *J. Comp. Physiol.* **178**, 797–802 (1996).

27. Brueggemann, L. I. & Sullivan, J. M. HEK293S cells have functional retinoid processing machinery. *J. Gen. Physiol.* **119**, 593–612 (2002).
28. Wong, S. K. G protein selectivity is regulated by multiple intracellular regions of GPCRs. *Neurosignals* **12**, 1–12 (2003).
29. Oh, E. J., Gover, T. D., Cordoba-Rodriguez, R. & Weinreich, D. Substance P evokes cation currents through TRP channels in HEK293 cells. *J. Neurophysiol.* **90**, 2069–2073 (2003).
30. Hardie, R. C. Regulation of TRP channels via lipid second messengers. *Annu. Rev. Physiol.* **65**, 735–759 (2003).
31. Melyan, Z., Tarrant, E. E., Bellingham, J., Lucas, R. J. & Hankins, M. W. Addition of human melanopsin renders mammalian cells photoreceptive. *Nature* doi:10.1038/nature03344 (this issue).
32. Panda, S. *et al.* Illumination of the melanopsin signaling pathway. *Science* **307**, 600–604 (2005).
33. Isoldi, M. C., Rollag, M. D., de Lauro Castrucci, A. M. & Provencio, I. Rhabdomic phototransduction initiated by the vertebrate photopigment melanopsin. *Proc. Natl Acad. Sci. USA* **102**(4), 1217–1221 (2005).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank V. Maine for technical support; M. Zhu for donating the HEK293-TRPC3 cells; T. Helton, K. Richard, D. Lipscombe, J. Bai and X. Wang for guidance with cell culture; E. Newman and D. O'Malley for advice on calcium imaging; and M. Rollag and J. McIlwain for discussions and comments on the manuscript. This work was supported by NIH grants to D.M.B. and I.P.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.B. (David_Berson@brown.edu).

Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN

Dennis M. Dacey¹, Hsi-Wen Liao², Beth B. Peterson¹, Farrel R. Robinson¹, Vivianne C. Smith³, Joel Pokorny³, King-Wai Yau² & Paul D. Gamlin⁴

¹University of Washington, Dept of Biological Structure and the Washington National Primate Research Center, Seattle, Washington 98195-7420, USA

²Departments of Neuroscience and Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205-2185, USA

³University of Chicago, Vision Science Laboratories, 940 East 57th Street, Chicago, Illinois 60637, USA

⁴University of Alabama at Birmingham, Vision Science Research Center, Birmingham, Alabama 35294-4390, USA

Human vision starts with the activation of rod photoreceptors in dim light and short (S)-, medium (M)-, and long (L)- wavelength-sensitive cone photoreceptors in daylight. Recently a parallel, non-rod, non-cone photoreceptive pathway, arising from a population of retinal ganglion cells, was discovered in nocturnal rodents¹. These ganglion cells express the putative photopigment melanopsin and by signalling gross changes in light intensity serve the subconscious, 'non-image-forming' functions of circadian photoentrainment and pupil constriction^{1–7}. Here we show an anatomically distinct population of 'giant' melanopsin-expressing ganglion cells in the primate retina that, in addition to being intrinsically photosensitive, are strongly activated by rods and cones, and display a rare, S-Off, (L + M)-On type of colour-opponent receptive field. The intrinsic, rod and (L + M) cone-derived light responses combine in these giant cells to signal irradiance over the full dynamic range of human vision. In accordance with cone-based colour opponency, the giant cells project to the lateral geniculate nucleus, the thalamic relay to primary visual cortex. Thus, in the diurnal trichromatic primate, 'non-image-forming' and conventional 'image-forming' retinal

pathways are merged, and the melanopsin-based signal might contribute to conscious visual perception.

There is evidence that a melanopsin-associated photodetective pathway exists in the diurnal human visual system^{8–12}, similar to the one found in the nocturnal rodent. However, the detailed anatomical and functional properties of a melanopsin pathway in primates, and its relationship to rod and cone circuits, are unknown. To identify the melanopsin-expressing cells in the primate, a polyclonal antibody derived from the conceptually translated, full-length complementary DNA for the human melanopsin protein was used to immunostain human and macaque retinæ. In flat mounts of the entire retina, the melanopsin antisera revealed a morphologically distinct population of ~3,000 retinal ganglion cells with completely stained cell bodies, dendritic trees and axons (Fig. 1a–c). With ~1.5 million ganglion cells in the human retina, the melanopsin-expressing cells comprise only 0.2% of the total. The melanopsin-expressing ganglion cell bodies were big, giving rise to the largest dendritic tree diameters of any primate retinal ganglion cell identified thus far¹³ (Fig. 1d–f). The long, sparsely branching dendrites produced an extensive meshwork of highly overlapping processes. Cell counts showed a shallow density gradient ranging from 3–5 cells mm⁻² over much of the retinal periphery to a peak of 20–25 cells mm⁻² in the parafoveal retina (Fig. 1g); in contrast, total ganglion cells reach a peak density of ~50,000 cells mm⁻². In the central retina, the extremely large dendritic trees of melanopsin-containing ganglion cells spiralled around the foveal pit to form an extensive plexus (Fig. 1e).

Melanopsin-containing dendrites were localized to two strata: the extreme inner and extreme outer borders of the inner plexiform layer (Fig. 1h). Individual cells are principally monostratified, creating two distinct subpopulations that send dendrites to either the inner or the outer stratum. About 60% of the melanopsin-expressing cells were outer-stratifying cells, and about 40% had cell bodies displaced to the inner nuclear layer

A novel cell-marking method was used in conjunction with the *in vitro* intact macaque retina¹⁴ to investigate the central targets and the visual physiology of the melanopsin-expressing cells. The tracer rhodamine dextran was injected into physiologically identified locations in the lateral geniculate nucleus (LGN) and pretectal olivary nucleus (PON), two major retinorecipient structures. After retrograde transport, the complete dendritic morphology of labelled ganglion cells was revealed in the *in vitro* retina by liberation of the sequestered rhodamine tracer into the cytoplasm upon light exposure¹³. Melanopsin immunostaining always colocalized with rhodamine in the cell body and dendritic tree of the tracer-labelled 'giant' cells (Fig. 2a–d). These distinctive 'giant', monostratified ganglion cells appear identical to those previously labelled in a similar retrograde fashion from both LGN and PON¹³.

The giant ganglion cells showed cone-driven input and an unexpected response to chromatic stimuli. At mid-photopic levels, a 550-nm light pulse evoked a sustained On response (Fig. 3a). Latencies to first spike were ~30–40 ms, typical of cone-mediated ganglion cell signals in primates (Fig. 3a inset). Surprisingly, a sustained On response was observed for both morphological cell

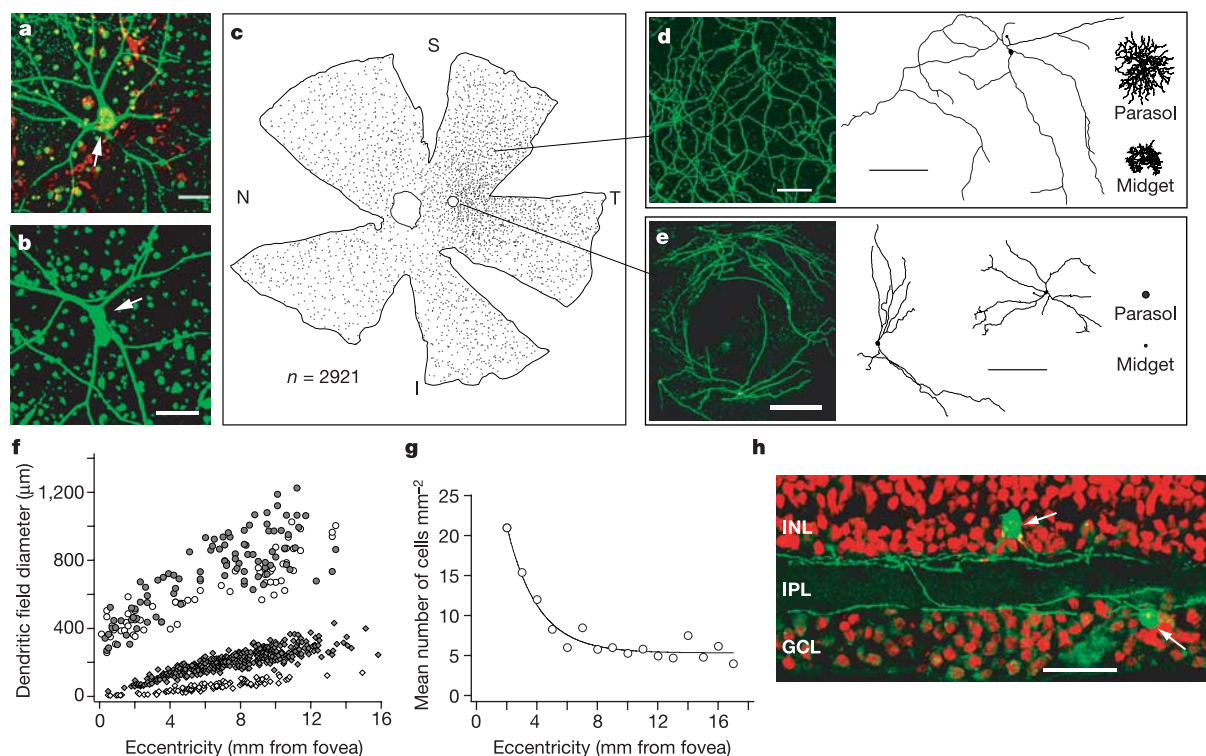


Figure 1 Morphology of melanopsin-immunoreactive cells. **a**, Human cell (arrow); propidium iodide red counterstain. Scale bar, 50 μm. **b**, Macaque cell (arrow). Scale bar, 50 μm. **c**, Macaque retina tracing; dots represent melanopsin cells. T, temporal retina; N, nasal retina; S, superior retina; I, inferior retina. **d**, Melanopsin cells in peripheral retina (left; scale bar, 100 μm). Tracing of a peripheral HRP-stained giant cell (right; scale bar, 200 μm). Parasol and midget cells (far right) are shown for comparison. **e**, Melanopsin cells encircling the fovea (left; scale bar, 200 μm). Tracings of two HRP-stained giant cells ~1–1.5 mm from the fovea (right; scale bar, 200 μm). Circles (far right) indicate size of

foveal parasol and midget cells. **f**, Dendritic field size of melanopsin cells versus eccentricity (inner cells, filled circles, $n = 93$; outer cells, open circles, $n = 63$). Parasol (filled diamonds, $n = 333$) and midget cells (open diamonds, $n = 93$) are shown for comparison. **g**, Mean cell density of melanopsin cells versus eccentricity (total 614 cells in 78 1 mm² samples). **h**, Dendritic arbours (green) of melanopsin cells (arrows) from stacked confocal images of 5 consecutive vertical sections (25 μm thick). The soma of the outer cell is displaced to the inner nuclear layer (INL). GCL, ganglion cell layer; IPL, inner plexiform layer. Scale bar, 50 μm.

populations (Fig. 3a), showing that the inner-versus-outer stratification, normally reflecting a division into On-centre versus Off-centre receptive field categories, does not apply here. Furthermore, these cells showed an unusual 'colour-opponent' receptive field in which an S cone-mediated Off response is antagonistic to an (L + M) cone-mediated On response (Fig. 3b). Cone-mediated receptive fields were large, approximating the dendritic tree diameter, and they showed spatially overlapping S-Off and (L + M)-On components, with little evidence of the strong inhibitory-surround typical of primate ganglion cells that project to the LGN (Fig. 3b, middle inset).

When the *in vitro* retina was maintained in total darkness for 10–20 min, light stimuli in the scotopic range elicited strong rod-driven responses (Fig. 3c). The response was a sustained On-type with the long latency to spike (~ 150 ms) (Fig. 3c, inset) and spectral tuning (peak at 502 nm, data not shown) characteristic of rod-driven input. The rod signal showed high photosensitivity, responding to quantal illuminances as low as 6–7 log quanta $\text{cm}^{-2} \text{s}^{-1}$ (or 4–5 log units below the threshold for a cone-mediated response), which is at or near the absolute threshold for human vision¹⁵.

Intrinsic photosensitivity was unmasked in the giant ganglion cells by pharmacologically blocking rod and cone transmission to the inner retina. Bath application of L-AP4 (DL-2-amino-4-phosphono-butyric acid) and CNQX (6-cyano-7-nitroquinoxaline-2,3-dione), which block both ionotropic and metabotropic

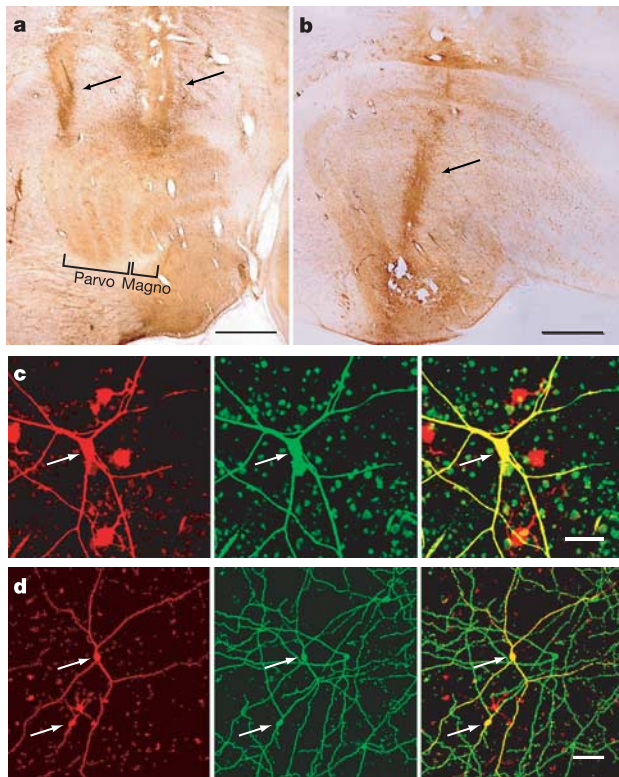


Figure 2 Retrograde tracer labelling from LGN and pretectum colocalize with melanopsin immunostain. **a**, **b**, Coronal sections through LGN showing HRP-stained tracer injection tracks (arrows). Injections in **a** are restricted to the parvocellular layers; the track in **b** extends to the magnocellular layers. Scale bars, 2 mm. **c**, Confocal images of retrograde rhodamine (red) labelling (left, arrow) from LGN injection and melanopsin (green) immunostaining (middle, arrow). Colocalization of labels appears yellow (right, arrow). Scale bar, 50 μm . **d**, Cells retrogradely labelled with rhodamine from pretectum injections (left, arrows) and melanopsin-immunostained (middle, arrows). Colocalization appears yellow (right, arrows). Scale bar, 100 μm .

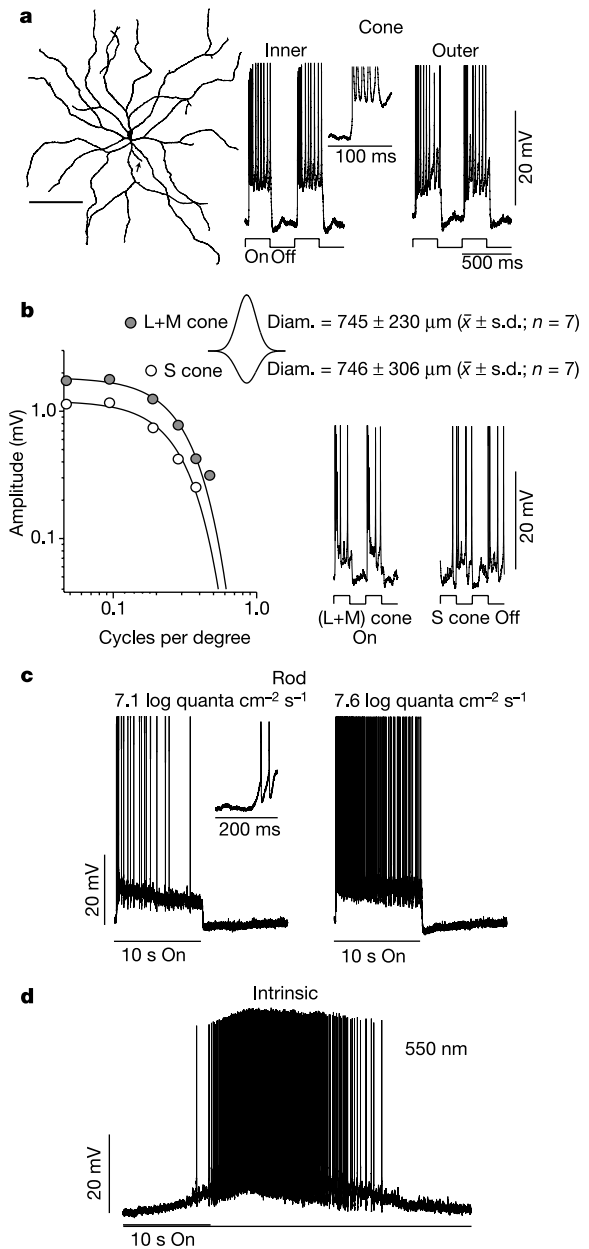


Figure 3 Giant cells show rod and colour-opponent inputs and are inherently photoreceptive. **a**, Tracing of a giant cell (arrow indicates axon; scale bar, 200 μm). The cell was recorded from and intracellularly filled with Neurobiotin in the *in vitro* retina. Voltage traces (right) show sustained On responses of an inner and an outer cell to a 2-Hz modulated 550-nm, full-field monochromatic light ($13.5 \log \text{quanta cm}^{-2} \text{s}^{-1}$) under photopic conditions. Inset: first 100 ms of voltage response of inner cell; response latency is 38 ms. Stimulus time indicated below voltage traces. **b**, The cell had an (L + M)-On, S-Off opponent receptive field. Plot (left) shows spatial frequency response to drifting gratings used to measure the receptive field; stimuli modulated the L + M cones (dark grey circles) or S cones (white circles) in isolation. Data were fitted with a difference-of-gaussians receptive field model (solid lines). Two-dimensional gaussian profile (middle) summarizes fits for 7 cells. Traces (far right) show responses to (L + M) and S cone-isolating stimuli, respectively. **c**, Pure rod-mediated responses elicited by a 550-nm monochromatic pulse at low scotopic levels. Inset: first 200 ms of voltage response at $7.1 \log \text{quanta cm}^{-2} \text{s}^{-1}$; response latency is 147 ms. **d**, L-AP4 and CNQX application block excitatory glutamatergic transmission, revealing a slow, sustained, inherent photoresponse (550-nm light; $13.5 \log \text{quanta cm}^{-2} \text{s}^{-1}$).

retinal glutamate receptors, completely eliminated the short-latency, cone-driven light response at photopic levels. However, long-duration light pulses still elicited a depolarizing voltage response that grew slowly and declined even more slowly after stimulus offset (Fig. 3d). This intrinsic photoresponse is much like that observed previously in melanopsin-expressing ganglion cells in rat under pharmacological blockade of all synaptic transmission^{1,6}. The long latency and sustained depolarization of the isolated intrinsic response was not an artefact created by drug application to the *in vitro* retina, because the latency to first spike varied systematically with stimulus strength, speeding up significantly at higher light levels (Fig. 4c). Moreover, in the drug-free condition, the fast cone-mediated response latency¹⁵ was observed in combination with the slow, sustained depolarization derived from the intrinsic response (Fig. 5d).

To investigate spectral tuning, we measured peak depolarization as a function of wavelength over a 3-log-unit illuminance range (Fig. 4a–c) to determine relative quantal sensitivity (Fig. 4d, e). These data were well-fitted by an A₁ visual pigment nomogram with a peak at 482 nm (Fig. 4f, mean $\pm$ s.d., 482 ± 1 nm; $n = 5$ cells), and were therefore distinct from primate rod, S, M and L cone pigments.

The regular and tonic spiking behaviour of the inherent photoresponse linearly encodes quantal illumination over a 3–4-log-unit range that encompasses most of daylight vision in the natural environment. The precise photon-counting capability in the macaque giant cells can be shown by counting the total number of spikes elicited by a long-duration light pulse, including those after light offset (Fig. 5a). Mean spike rate also increases regularly with increasing irradiance (Fig. 5b). Thus, the inherent photoresponse, when isolated, appears to transmit a signal that best measures total retinal irradiance when accumulated over long time periods; this property could serve as a neural memory of long-term light history to be used by the circadian system¹⁶.

How does the intrinsic photoresponse of the giant ganglion cells interact with the rod- and/or cone-mediated signals under physiological conditions? Irradiance coding is present in the dark-adapted, pure rod response (Fig. 5b) but is not normally associated with the cone signal¹⁷. Without pharmacological blockade of the cone input, it was possible to observe a summation of short-latency depolarizing cone response and slow intrinsic response over most of the photopic range at 470 nm (Fig. 5b, d). On the surface, the inherent response appears to simply elevate the cone-mediated spike rate and sustain the firing beyond the duration of light stimulus (Fig. 5d, e). In actuality, however, the inherent response serves to compensate for the relative transience of the cone signal. This interaction can be appreciated by comparing the (L + M) cone-mediated On response elicited by a long-wavelength light with the mixed cone/intrinsic response elicited by a short-wavelength light at about the same quantal level (Fig. 5f, g). The (L + M) cone-mediated response is transient at threshold (Fig. 5f, left) and, even at the highest photopic levels used in this study (Fig. 5f, right), declines rapidly during a maintained light step. In contrast, when a short-wavelength light is used, the depolarizing, intrinsic response is added, making the response at threshold more sustained and apparently masking any transient inhibition from S cones (Fig. 5g, left). At higher light levels the cone + intrinsic response shows characteristic continued discharge after light offset (Fig. 5g, right).

We have shown that some basic properties of the melanopsin pathway, such as the dynamics and spectral tuning of the inherent light response and a projection to the brainstem pupillomotor nucleus, are conserved from rodents to primates. However, our findings reveal a fundamental contribution of rod/cone signals to this circuit in the diurnal primate and at the same time provide the first evidence of a broader role for this pathway in higher visual processing. By combining the rod, cone and inherent photo-

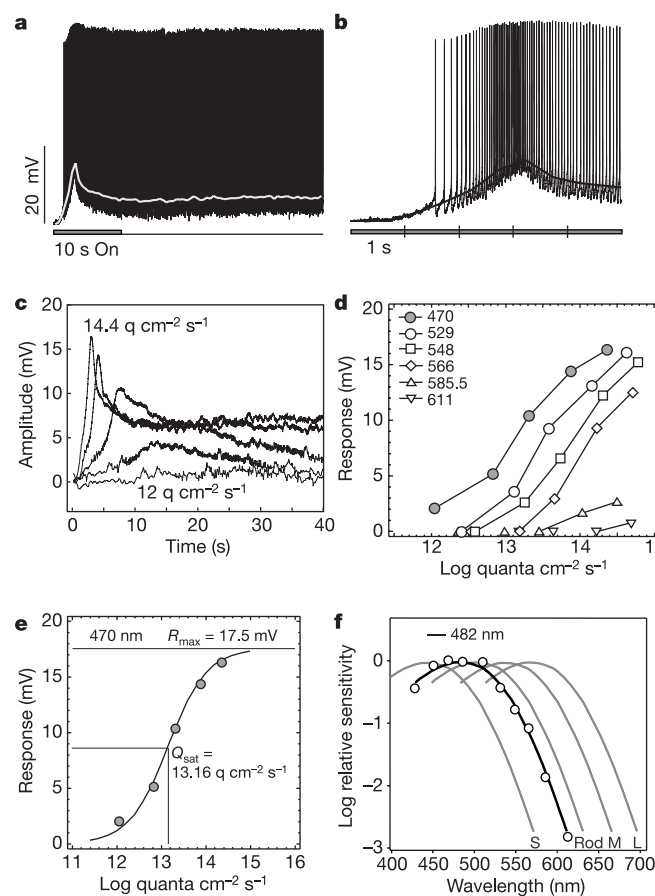


Figure 4 Spectral sensitivity of the giant cell's inherent light response. **a**, Response to a 470-nm pulse at $14.35 \log \text{ quanta cm}^{-2} \text{ s}^{-1}$ showed a peak of 40 impulses s^{-1} near 3 s, and continued firing for 30 s after the end of the light stimulus. White line shows membrane potential values averaged over 0.5 s sliding time windows. **b**, First 5 s of the response shown in **a**; averaged values pass near the spike initiation event. **c**, Averaged responses to 470-nm stimulus over a 12–14.5 $\log \text{ quanta cm}^{-2} \text{ s}^{-1}$ illuminance range; peak response was 16 mV for the brightest stimulus. **d**, Plot of averaged response peaks for 6 wavelengths between 470 and 611 nm as a function of quantal illumination. **e**, Michaelis–Menten equation fitted to the 470-nm data; responsivity was assessed for 10 wavelengths. **f**, Responsivity scaled for best fit to an A₁-based photopigment nomogram of 482 nm (solid line).

responses, the giant ganglion cell pathway can by itself convey signals from all receptor classes known to drive the circadian and pupillomotor systems^{2,7}. Through chromatic opponency, the cone circuitry may have originally evolved to signal the large spectral changes at dawn and dusk to the circadian pathway in order to more precisely set the biological clock to the solar day¹⁸. Along the primary visual pathway, neurons with S-Off and S-On opponent receptive fields have been previously identified in the LGN¹⁹ and primary visual cortex²⁰, and are basic components of psychophysical models of human colour vision²¹. S-On signals originate from at least two novel ganglion cell populations^{13,14}, but the origin of an S-Off signal has remained uncertain²². The giant cells demonstrate that S-Off, like S-On opponency, can also derive from a distinct ganglion cell population. Reciprocally, the sustained, irradiance-coding signal mediated by the intrinsic light response would presumably also reach primary visual cortex via the LGN. Over 30 years ago, Horace Barlow first called attention to a few exceptional 'luminance coding' units in the cat's retina, in which spike rate increased monotonically with increasing irradiance²³. Similar

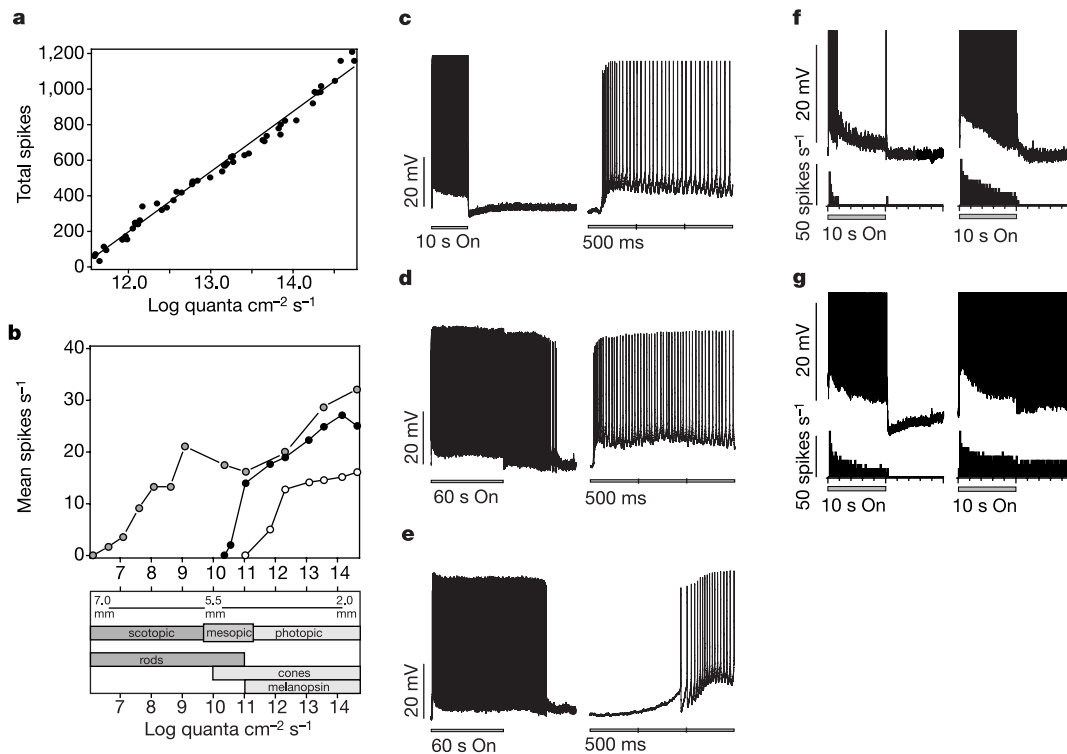


Figure 5 Irradiance coding and visual sensitivity range for the giant cell. **a**, Total spikes following a 10-s pulse for each wavelength used to measure spectral tuning (see Fig. 4), plotted versus log quanta $\text{cm}^{-2} \text{s}^{-1}$, weighted by the probability of absorption calculated from the Op^{482} nomogram (linear regression fit, $r = 0.99$, slope = 339 spikes per log unit). **b**, Response of a giant cell to a monochromatic light step (470 nm) as a function of retinal illuminance for 3 stimulus conditions: dark-adapted (grey circles), 10-s pulse after 20 min total darkness; light-adapted (black circles), 60-s pulse immediately after several min at high photopic levels (>14 quanta $\text{cm}^{-2} \text{s}^{-1}$); isolated intrinsic (open circles), 60-s pulse after light adaptation in the presence of bath-perfused L-AP4 and CNQX. Boxed area below plot shows melanopsin-associated (melanopsin) and rod and cone response ranges in relation to scotopic, mesopic, and photopic ranges of human vision and pupil diameter¹⁵. **c–e**, Intracellularly recorded light responses under the 3 conditions shown in **b**. **c**, Rod-mediated (dark-adapted) response at 8.1 log quanta $\text{cm}^{-2} \text{s}^{-1}$; latency to first spike, 147 ms. **d**, Cone-mediated (light-adapted) response at 13.5 log quanta $\text{cm}^{-2} \text{s}^{-1}$; latency, 36 ms. Prolonged spiking due to inherent photoreponse is seen at the end of the 60 s stimulus pulse. **e**, Inherent (isolated intrinsic) response at 13.5 log quanta $\text{cm}^{-2} \text{s}^{-1}$ after block of cone input; latency, 903 ms. **f**, Pure cone-mediated responses to 610-nm light pulse at both low (left; 12 log quanta $\text{cm}^{-2} \text{s}^{-1}$) and high (right; 15.2 log quanta $\text{cm}^{-2} \text{s}^{-1}$) photopic levels; post-stimulus time-spike histograms are below voltage traces. **g**, Summed cone and intrinsic response to 470-nm pulse at low (left; 11 log quanta $\text{cm}^{-2} \text{s}^{-1}$) and high (right; 14.6 log quanta $\text{cm}^{-2} \text{s}^{-1}$) photopic levels.

irradiance-coding units have been recorded in macaque LGN²⁴ and primary visual cortex²⁵, with a maintained discharge rate set by the overall level of diffuse illumination. The origin and spectral signature of this higher signal has yet to be determined, but it might arise from the giant cells described here and play a role in the conscious perception of brightness^{26,27} □

Methods

All experimental procedures were approved by the Institutional Animal Care and Use Committees at either the University of Washington or the University of Alabama.

Antibody preparation

A peptide consisting of 19 amino acid residues, MNPPSGPRVPPSPTEPSC, from the N terminus of the conceptually translated human melanopsin protein (NCBI accession number AAF24978) was synthesized (Princeton BioMolecules), with an additional lysine residue at the amino end for crosslinking. Purified peptide was crosslinked to thyroglobulin using glutaraldehyde, then dialysed in PBS buffer. The peptide conjugate was used to immunize rabbits (Covance); the resulting antiserum was purified using peptide-bovine serum albumin (BSA)-packed affinity column.

Melanopsin immunostaining

Adult human and macaque retinæ were fixed flat in 4% paraformaldehyde for 2 h, then washed in 0.1 M phosphate buffer (pH 7.4). Whole retinæ were blocked in 0.1% Triton-X100 (Sigma) and 0.5% BSA in PBS for 3 h. The melanopsin antibody (1:100) was added and retinæ were incubated for 4 days at 4 °C. After washing, macaque retinæ were incubated in either biotinylated goat anti-rabbit antibody (BA-1000, Vector Labs) (8 animals) at 1:100 in 0.1% Triton-X100 and PBS for 48 h at 4 °C, or in Alexa Fluor 488 goat anti-rabbit antibody (A-11034, Molecular Probes) (1 animal) at 1:100 in PBS for 48 h

at 4 °C. Human retinæ were processed with the Alexa Fluor secondary antibody only. Retinæ treated with biotinylated antibody were incubated in 0.1% Triton-X100 containing the Vector avidin-biotin-HRP complex (ABC Elite kit, Vector Labs) for 1–2 days at 4 °C, rinsed in phosphate buffer, and processed for HRP histochemistry by incubation in diaminobenzidine (DAB, 0.1% in 0.01 M phosphate buffer, pH 7.4) for 5 min, followed by addition of H_2O_2 (0.03%) and further incubation for 3–4 min. Frozen vertical sections (25 μm) of macaque retina were reacted using the same protocol except that primary antibody incubation was 10–12 h, secondary antibody incubation was 7–8 h, and avidin-biotin-HRP incubation was ~4–5 h. Sections were counterstained with the nuclear dye propidium iodide.

Cell counts

Melanopsin-immunostained cells were counted by taking sequential confocal images of all labelled cells, then 'splicing' the images together to reconstruct the entire retina. A grid was placed over the image of the reconstructed retina, centred on the fovea, and a small dot was placed over each fluorescent cell. The cells in each mm^2 of the grid were counted and the total number of labelled cells in the retina mapped.

Rhodamine-melanopsin colocalization

In two monkeys, biotinylated rhodamine dextran was used to retrogradely tracer-label retinal ganglion cells projecting to the LGN and pretectum, followed by melanopsin immunostaining of the retinæ. Animals were anesthetized and prepared for recording in an aseptic surgery as described elsewhere¹³. In brief, the position of the LGN or pretectum was determined stereotactically and evaluated by mapping extracellular responses to flashes of light. In each targeted area, ~0.5 μl injections were made of 10% biotinylated dextran-conjugated tetramethylrhodamine 3,000 MW (micro ruby, #D-7162, Molecular Probes) in sterile saline. After 4–7 days, the animal was deeply anesthetized, the eyes removed, and retinæ prepared for the *in vitro* experiment. Animals were perfused through the heart with 800 ml warm (~37 °C) normal saline followed by 4 litres cold (~4 °C) 4% paraformaldehyde. Frozen brain sections were processed as described above for the retina,

to reveal biotin labelling at the injection sites. Rhodamine-labelled cells in the retina were visualized *in vitro* with a green filter block (excitation filter 545 nm, barrier filter 590 nm). The fluorescent label, initially confined to small bright spots in the soma and proximal dendrites, diffused throughout the entire dendritic tree after brief light exposure¹³. After the *in vitro* experiment, retinas were fixed and processed for melanopsin immunoreactivity as described above. Confocal images of rhodamine (red) and Alexa Fluor 488 (green) labelled cells were used to demonstrate colocalization.

In vitro preparation and intracellular recording

The *in vitro* retina preparation and recording have been described previously¹⁴. Eyes were removed from deeply anesthetized animals and the retina was isolated from the vitreous and sclera in oxygenated Ames' Medium (Sigma). The retina-RPE-choroid was placed flat, vitreal surface up, in a superfusion chamber mounted on the stage of a light microscope. Rhodamine-labelled cells were visualized as described above. Intracellular recording of targeted cells was done using high-impedance (~300–450 MΩ) glass micropipettes filled with 2–3% Neurobiotin (Vector Labs) and 1–2% pyranine (Molecular Probes) in 1.0 M potassium acetate. Voltage responses were amplified (Axoclamp, Axon Instruments) and digitized at 10 kHz.

Light stimulation

Square-wave pulses and sinusoidally-modulated stimuli were created using a VSG3-series stimulus generator (Cambridge Research Systems), which produced video input to a digital light projector (Vista Pro Plus, Christie Digital Systems). The projector output was relayed to the microscope camera port and focused on the retina by a microscope objective. A detailed description of the stimulator has been published²⁸. Spectral opponency was evaluated with stimuli that modulated S-cones or (L + M) cones in isolation²⁹. Drifting sinusoidal gratings (2 Hz temporal frequency) that modulated either the S-cones or the (L + M) cones in isolation but varied in spatial frequency were used to determine receptive field structure. Response amplitudes to a cone-isolating grating series were used to determine the parameters of a difference-of-gaussians receptive field model³⁰.

L-AP4 and CNQX application

A solution of 100 μM L-AP4 (Sigma) and 75 μM CNQX (Tocris) in oxygenated Ames' medium was applied to the retina. All cone-driven responses to temporally modulated square-wave stimuli (0.5–10 Hz) disappeared within 2–3 min of application. Drugs were superfused for up to 1–12 h during recordings of inherent light responses of giant cells ($n = 27$). Cone-mediated responses recovered within a few minutes of drug washout in all cases.

Spectral tuning

After pharmacological isolation of rod- and cone-mediated responses, a Varispec liquid crystal tuneable imaging filter inserted in the light path between projector and microscope was used to select narrow-band stimuli (15–20 half-bandwidth) at ten wavelengths (430–610 nm in 20-nm steps). Out-of-band light was minimized by selecting the suitable digital light projector primary for each of the wavelengths during stimulus presentation, and by setting the Varispec wavelength to 700 nm between stimulus presentations. The light stimulus was a 10-s pulse followed by 30 s recovery in the dark. At each wavelength, ganglion cell responses were measured at 0.5 log unit illuminance intervals from 10.5–14.5 log quanta $\text{cm}^{-2} \text{s}^{-1}$. Peak response was plotted as a function of quantal illumination; responsivity was assessed for each wavelength and fit to a visual pigment A_1 nomogram. Nomograms were scaled in 1-nm steps.

Received 21 September 2004; accepted 21 January 2005; doi:10.1038/nature03387.

- Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
- Hattar, S. *et al.* Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. *Nature* **424**, 75–81 (2003).
- Lucas, R. J. *et al.* Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).
- Panda, S. *et al.* Melanopsin (Opn4) requirement for normal light-induced circadian phase shifting. *Science* **298**, 2213–2216 (2002).
- Ruby, N. F. *et al.* Role of melanopsin in circadian responses to light. *Science* **298**, 2211–2213 (2002).
- Hattar, S., Liao, H. W., Takao, M., Berson, D. M. & Yau, K. W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Panda, S. *et al.* Melanopsin is required for non-image-forming photic responses in blind mice. *Science* **301**, 525–527 (2003).
- Thapan, K., Arendt, J. & Skene, D. J. An action spectrum for melatonin suppression: evidence for a novel non-rod, non-cone photoreceptor system in humans. *J. Physiol. (Lond.)* **535**, 261–267 (2001).
- Provencio, I. *et al.* A novel human opsin in the inner retina. *J. Neurosci.* **20**, 600–605 (2000).
- Brainard, G. C. *et al.* Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *J. Neurosci.* **21**, 6405–6412 (2001).
- Hankins, M. W. & Lucas, R. J. The primary visual pathway in humans is regulated according to long-term light exposure through the action of a nonclassical photopigment. *Curr. Biol.* **12**, 191–198 (2002).
- Hannibal, J. *et al.* Melanopsin is expressed in PACAP-containing retinal ganglion cells of the human retinohypothalamic tract. *Invest. Ophthalmol. Vis. Sci.* **45**, 4202–4209 (2004).
- Dacey, D., Peterson, B., Robinson, F. & Gamlin, P. Fireworks in the primate retina: *in vitro* photodynamics reveals diverse LGN-projecting ganglion cell types. *Neuron* **37**, 15–27 (2003).
- Dacey, D. M. & Lee, B. B. The blue-ON opponent pathway in primate retina originates from a distinct bistrifid ganglion cell type. *Nature* **367**, 731–735 (1994).
- Hood, D. C. & Finkelstein, M. A. In *Handbook of Perception and Human Performance* Vol. 1 (eds Boff, K. R., Kaufman, L. & Thomas, J. P.) Ch. 5, 1–66 (John Wiley and Sons, New York, 1986).
- Takahashi, J. S., DeCoursey, P. J., Bauman, L. & Menaker, M. Spectral sensitivity of a novel

- photoreceptive system mediating entrainment of mammalian circadian rhythms. *Nature* **308**, 186–188 (1984).
- Fain, G. L., Matthews, H. R., Cornwall, M. C. & Koutalos, Y. Adaptation in vertebrate photoreceptors. *Physiol. Rev.* **81**, 117–151 (2001).
- Mollon, J. D. & Jordan, G. Eine evolutionäre Interpretation des menschlichen Farbensehens. *Die Farbe* **35/36**, 139–170 (1989).
- Valberg, A., Lee, B. B. & Tigwell, D. A. Neurones with strong inhibitory s-cone inputs in the macaque lateral geniculate nucleus. *Vision Res.* **26**, 1061–1064 (1986).
- Cottaris, N. & DeValois, R. Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature* **395**, 896–900 (1998).
- Krauskopf, J., Williams, D. R. & Heeley, D. W. Cardinal directions of color space. *Vision Res.* **22**, 1123–1131 (1982).
- Klug, K., Herr, S., Ngo, I. T., Sterling, P. & Schein, S. Macaque retina contains an S-cone OFF midget pathway. *J. Neurosci.* **23**, 9881–9887 (2003).
- Barlow, H. B. & Levick, W. R. Changes in the maintained discharge with adaptation level in the cat retina. *J. Physiol. (Lond.)* **202**, 699–718 (1969).
- Marrocco, R. T. Possible neural basis of brightness magnitude estimations. *Brain Res.* **86**, 128–133 (1975).
- Kayama, Y., Riso, R. R., Bartlett, J. R. & Doty, R. W. Luxotonic responses of units in macaque striate cortex. *J. Neurophysiol.* **42**, 1495–1517 (1979).
- Kinoshita, M. & Komatsu, H. Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *J. Neurophysiol.* **86**, 2559–2570 (2001).
- Barlow, R. B. Jr & Verrillo, R. T. Brightness sensation in a ganzfeld. *Vision Res.* **16**, 1291–1297 (1976).
- Packer, O. *et al.* Characterization and use of a digital light projector for vision research. *Vision Res.* **41**, 427–439 (2001).
- Diller, L. *et al.* L and M cone contributions to the midget and parasol ganglion cell receptive fields of macaque monkey retina. *J. Neurosci.* **24**, 1079–1088 (2004).
- Enroth-Cugell, C., Robson, J. G., Schweitzer-Tong, D. E. & Watson, A. B. Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *J. Physiol. (Lond.)* **341**, 279–307 (1983).

Acknowledgements We would like to thank C. Curcio and the Age-Related Maculopathy Histopathology Laboratory (supported by the International Retinal Research Foundation, the National Eye Institute and the Vision Science Research Center), University of Alabama at Birmingham for the human retinas used in the immunohistochemical studies. Macaque retinas were provided by the Tissue Distribution program of the National Primate Research Center at the University of Washington. We thank O. Packer and T. Haun for technical assistance. Supported by US National Eye Institute grants to D.M.D., J.P., K.-W.Y., H.W.-L. and F.R.R., Vision Research Center Core grants to D.M.D. and P.D.G., an Alabama EyeSight Foundation award to P.D.G. and a Retina Research Foundation Paul Kayser Award to D.M.D.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.M.D. (dmd@u.washington.edu).

DRP-1-mediated mitochondrial fragmentation during EGL-1-induced cell death in *C. elegans*

Ravi Jagasia^{1,2,3}, Phillip Grote³, Benedikt Westermann^{2,4} & Barbara Conrad³

¹Max Planck Institute of Neurobiology, Am Klopferspitz 18a, D-82152 Planegg-Martinsried, Germany

²Institut für Physiologische Chemie, Universität München, D-81377 München, Germany

³Dartmouth Medical School, Department of Genetics, 7400 Remsen, Hanover, New Hampshire 03755, USA

⁴Zellbiologie, Universität Bayreuth, D-95440 Bayreuth, Germany

Genetic analyses in *Caenorhabditis elegans* have been instrumental in the elucidation of the central cell-death machinery, which is conserved from *C. elegans* to mammals^{1,2}. One possible difference that has emerged is the role of mitochondria. By releasing cytochrome *c*, mitochondria are involved in the activation of caspases in mammals^{3,4}. However, there has previously been no evidence that mitochondria are involved in caspase activation in *C. elegans*. Here we show that mitochondria fragment in cells that normally undergo programmed cell death during *C. elegans* development. Mitochondrial fragmentation is

induced by the BH3-only protein EGL-1 and can be blocked by mutations in the *bcl-2*-like gene *ced-9*, indicating that members of the Bcl-2 family might function in the regulation of mitochondrial fragmentation in apoptotic cells. Mitochondrial fragmentation is independent of CED-4/Apaf-1 and CED-3/caspase, indicating that it occurs before or simultaneously with their activation. Furthermore, DRP-1/dynamin-related protein, a key component of the mitochondrial fission machinery, is required and sufficient to induce mitochondrial fragmentation and programmed cell death during *C. elegans* development. These results assign an important role to mitochondria in the cell-death pathway in *C. elegans*.

Of the 1,090 somatic cells formed during the development of *C. elegans*, 131 undergo programmed cell death^{5,6}. The death of these cells is due to the activities of four genes, namely *egl-1* (*egl*, for egg-laying defective), *ced-9* (*ced*, for cell-death defective), *ced-4* and *ced-3*, which encode a pro-apoptotic BH3-only protein (*egl-1*), an anti-apoptotic Bcl-2-like protein (*ced-9*), an Apaf-1-like adaptor protein (*ced-4*) and a pro-caspase (*ced-3*)¹. In cells destined to survive, the mitochondria-localized CED-9 protein binds to CED-4, thereby blocking CED-4's ability to activate the caspase CED-3 (refs 1, 7). In cells destined to die, *egl-1* transcription is activated⁸. Binding of the EGL-1 protein to CED-9 causes CED-4 to relocate to perinuclear membranes resulting in the activation of CED-3 and CED-3-dependent killing⁷.

Recently, the mitochondrial network in cultured mammalian cells has been shown to undergo fragmentation during apoptosis^{9,10}, a process that might be involved in the release of cytochrome *c* from mitochondria^{11,12}. To investigate whether mitochondrial morphology changes occur in cells destined to die in *C. elegans*, we expressed a mitochondrial matrix-targeted green fluorescent protein (mitoGFP) under the control of the *egl-1* promoter (*P_{egl-1}mitogfp*) in wild-type embryos. The transcriptional activation of *egl-1* is an early event during programmed cell death⁸. The appearance of a GFP signal therefore marks the onset of the apoptotic process and specifically labels apoptotic cells. Mitochondria in apoptotic and non-apoptotic cells were stained with the fluorescent dye rhodamine B hexyl ester and examined by confocal time-lapse microscopy. After activation of the *egl-1* promoter, newly synthesized mitoGFP accumulated in the cytosol before being imported into mitochondria. Rhodamine-stained mitochondria appeared as tubular networks during the initial phase of programmed cell death and were indistinguishable from mitochondria in non-apoptotic—that is, GFP-negative—cells (Fig. 1a, wild-type, 0 min). Within a short period, the morphology of mitochondria changed in apoptotic cells. The mitochondrial network started to break down (8 min 52 s), eventually resulting in a few clusters of mitochondrial fragments located at the periphery of the cells (16 min 8 s). These changes in mitochondrial morphology were apparent before the apoptotic cells turned into characteristic 'refractile corpses' when observed by differential interference contrast microscopy (DIC)⁵. Mitochondrial fragmentation is therefore an early event in the apoptotic process.

To determine whether mitochondrial fragmentation in apoptotic cells is dependent on the initiation of apoptosis by EGL-1, we analysed the morphology of mitochondria in mitoGFP-positive cells in animals homozygous for the *egl-1* loss-of-function mutation *n1084 n3082*, which blocks programmed cell death during *C. elegans* development¹³. We found that, in this mutant background, mitochondria in GFP-positive cells retained their tubular network structure even when being observed for extended periods after the activation of *P_{egl-1}mitogfp* expression (Fig. 1a, *egl-1*). As the *egl-1(n1084 n3082)* mutation does not affect mitochondrial morphology in general (Supplementary Fig. S1), we conclude that its effect on mitochondrial fragmentation is specific for apoptotic cells.

Ectopic *egl-1* expression during embryogenesis results in ectopic

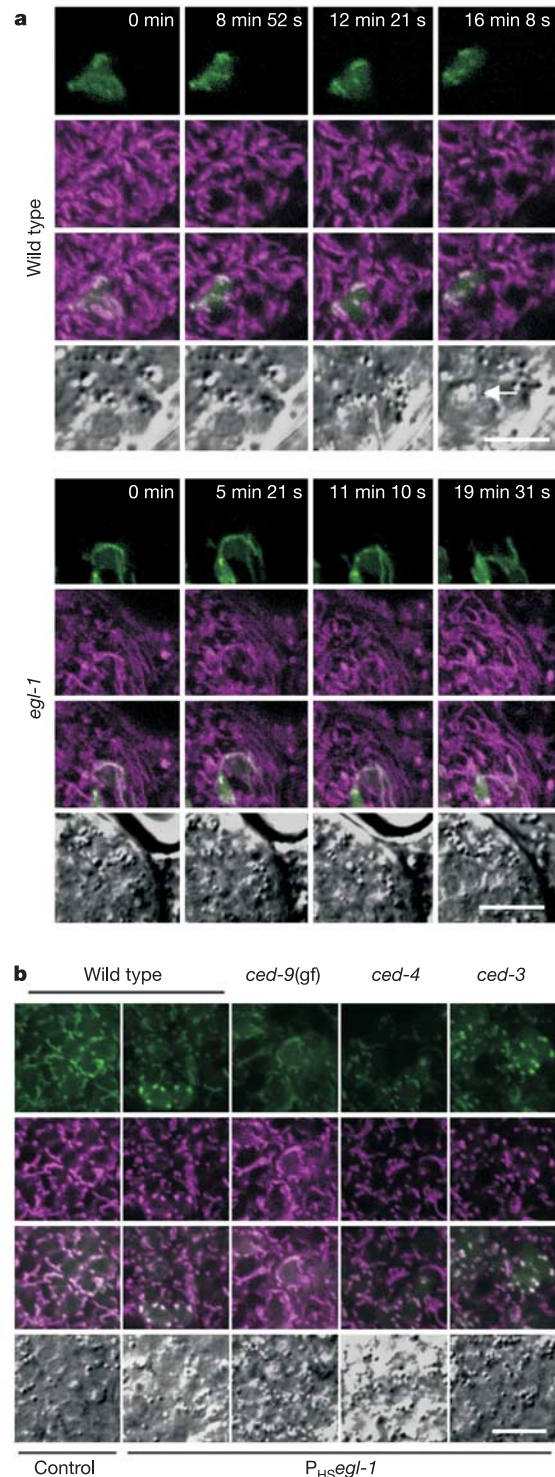


Figure 1 Mitochondrial morphology in cells undergoing programmed cell death in *C. elegans*. **a**, Representative time-lapse series of confocal mitoGFP, rhodamine, mitoGFP/rhodamine overlay and DIC images (from top to bottom) of wild-type and *egl-1(n1084 n3082)* animals carrying a stable *P_{egl-1}mitogfp* transgene (*bcl-49*). Images represent maximum intensity projections from four consecutive confocal image planes (0.5 μm). Scale bars, 4 μm. The arrow points to a refractile corpse. **b**, Representative confocal images of mitoGFP, rhodamine, mitoGFP/rhodamine overlay and DIC (from top to bottom) of transgenic wild-type, *ced-9(n1950gf)*, *ced-4(n1162)* and *ced-3(n717)* animals carrying a *P_{HS}mitogfp* transgene alone (wild-type, control, left column) or in combination with a *P_{HS}egl-1* transgene (all other columns). After induction of the transgenes, embryos were imaged at the comma to 1½-fold stage of embryonic development. Images represent single confocal image planes. Scale bar, 8 μm.

programmed cell death, which is blocked by a gain-of-function mutation in the *ced-9* gene (*n1950gf*) or by loss-of-function mutations in either the *ced-4* (*n1162*) or the *ced-3* gene (*n717*)¹³. We found that ectopic *egl-1* expression under the control of a heat-inducible promoter ($P_{HSEgl-1}$) induced mitochondrial changes reminiscent of the mitochondrial fragmentation observed in cells normally destined to undergo programmed cell death (Fig. 1b, compare Fig. 1a). EGL-1-induced mitochondrial fragmentation was blocked by *ced-9*(*n1950gf*) but not by *ced-4*(*n1162*) or *ced-3*(*n717*) (Fig. 1b). These three mutations do not affect mitochondrial morphology on their own, as determined by rhodamine

staining of embryos (Supplementary Fig. S1). Thus, EGL-1-induced mitochondrial fragmentation in apoptotic cells is blocked by *ced-9*(*n1950gf*); however, it does not require *ced-3* or *ced-4*.

It has recently been reported that overexpression of the gene *icd-1* (*icd*, for inhibitor of cell death) of *C. elegans*, which encodes the β -subunit of the nascent polypeptide-associated complex, blocks programmed cell death during *C. elegans* development¹⁴. We therefore sought to determine whether ectopic *icd-1* expression blocks *egl-1*-induced programmed cell death and mitochondrial fragmentation. However, using standard assays, we failed to observe any anti-apoptotic activity for *icd-1* (Supplementary Table S1, Supplementary

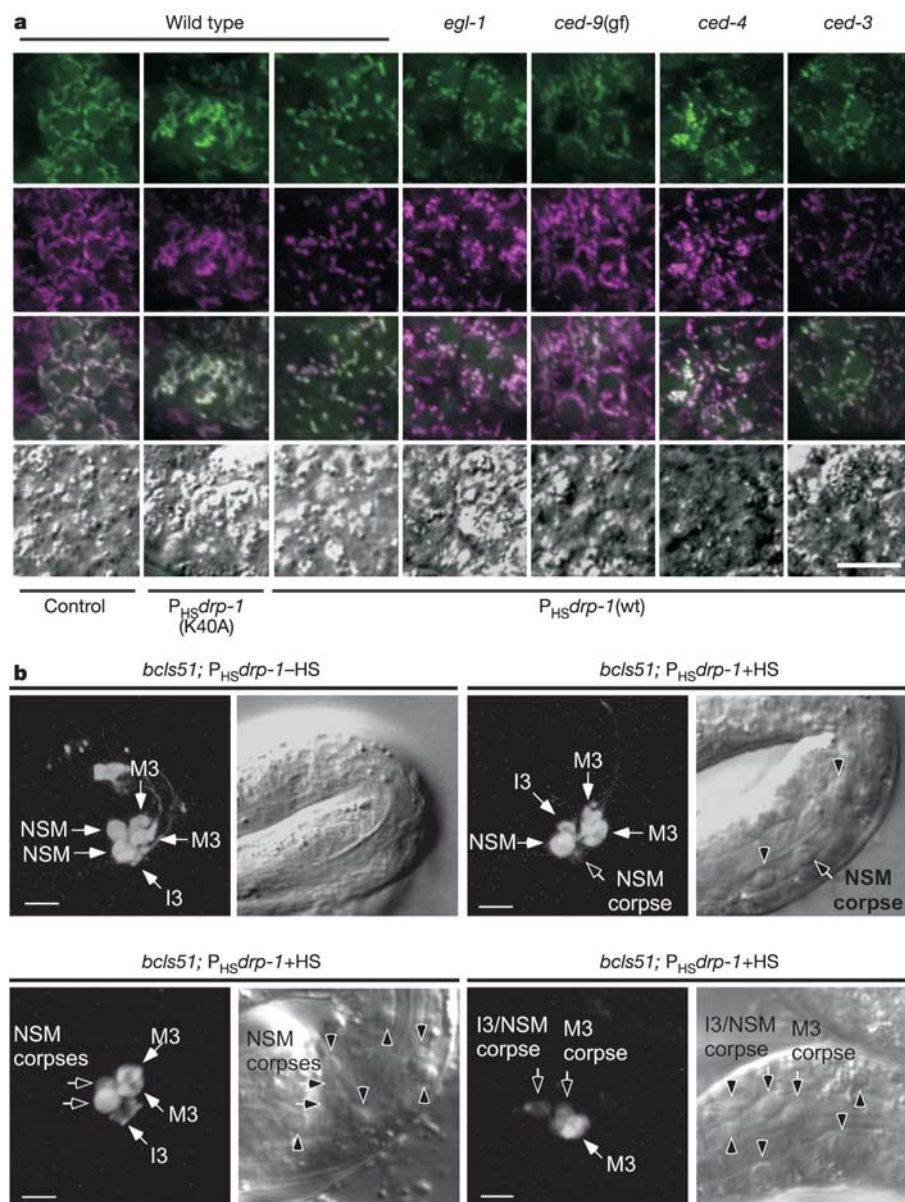


Figure 2 Expression of *drp-1*(K40A) or *drp-1*(wt). **a**, Mitochondrial morphology in embryos expressing *drp-1*(K40A) or *drp-1*(wt). Representative confocal images of mitoGFP, rhodamine, mitoGFP/rhodamine overlay and DIC (from top to bottom) of wild-type animals carrying a $P_{HSEgl-1}$ transgene alone (control, first column), in combination with a $P_{HSEgl-1}$ (K40A) transgene (*drp-1*(K40A), second column) or in combination with a $P_{HSEgl-1}$ (wt) transgene (*drp-1*(wt), third column), and *egl-1*(*n1084 n3082*), *ced-9*(*n1950gf*), *ced-4*(*n1162*), or *ced-3*(*n717*) animals carrying $P_{HSEgl-1}$ and $P_{HSEgl-1}$ (wt) (other columns, as indicated). After induction of the transgenes, embryos were imaged at the comma to 1½-fold stage of embryonic development. Images represent single confocal

image planes. Scale bar, 8 μ m. **b**, Expression of *drp-1*(wt) in embryos causes ectopic cell death. Representative confocal dsRed and DIC images of wild-type animals carrying a stable integration of $P_{HSEgl-1}$ -*dsRed* (*bcls51*), and a $P_{HSEgl-1}$ (wt) and $P_{HSEgl-1}$ transgene at the 3½–4-fold stage of embryonic development, not subjected to heat shock (–HS) or subjected to heat shock (+HS). Heat shock was applied as described in Methods except that embryos were collected over a period of 90 min and, 4 h later, were heat shocked for 60 min. Scale bars, 5 μ m. White arrows indicate living dsRed-positive NSMs, M3s or I3s; black arrows indicate refractile dsRed-positive NSM, M3 or I3 corpses. Black arrowheads indicate additional, unidentified refractile corpses.

Table 1 Expression of *drp-1(K40A)* with inducible promoter blocks cell death

Line	Without heat shock			With heat shock		
	Average number of extra cells	s.d.	Range	Average number of extra cells	s.d.	Range
Control (1)	0.1	0.3	0–1	0.2	0.2	0–1
Control (2)	0	0	0	0.1	0.1	0–1
Control (3)	0.1	0.3	0–1	0.1	0.1	0–1
<i>P_{HS}drp-1(K40A)</i> (1)	0.4	0.7	0–2	2.9	1.5	1–5
<i>P_{HS}drp-1(K40A)</i> (2)	0.5	0.8	0–2	2.2	1.7	0–6
<i>P_{HS}drp-1(K40A)</i> (3)	0.9	1.5	0–5	2.8	1.9	0–7
<i>P_{HS}drp-1(K40A)</i> (4)	1.0	1.0	0–3	2.7	1.4	0–4

Transgenes were activated and the numbers of extra cells in the anterior pharynx were determined as described in Methods. s.d., standard deviation. The number of animals counted was $n=10-15$. The complete genotypes of the animals were (from top to bottom): *unc-76(e911)*; *P_{HS}mitogfp* lines 1, 2 and 3, *unc-76(e911)*; *P_{HS}mitogfp+P_{HS}drp-1(K40A)* lines 1, 2, 3 and 4.

Figs S2 and S3). The reason for this is currently unclear.

Dynamin-related GTP-binding proteins have been shown to be required for the fission of mitochondria in yeast, *C. elegans* and mammals^{15,16}. It has also been shown that the expression of a dominant-negative form of the human dynamin-related protein DRP1 prevents mitochondrial fragmentation in cultured mammalian cells exposed to a cell-death stimulus and can delay their subsequent death¹¹. To determine whether mitochondrial fragmentation is required for programmed cell death during *C. elegans* development, we decreased the rate of mitochondrial fission by reducing the activity of *C. elegans* DRP-1 (*drp*, for dynamin-related protein) using a dominantly interfering mutation in the *drp-1* gene, *drp-1(K40A)*¹⁶. The expression of *drp-1(K40A)* in muscle cells of adult *C. elegans* causes mitochondria to lose their normal structure and to form large 'blebs'¹⁶. The expression of *drp-1(K40A)* during embryogenesis using the heat-inducible promoter *P_{HS}drp-1(K40A)* similarly resulted in the formation of aberrant mitochondria (Fig. 2a, *P_{HS}drp-1(K40A)*), which resemble interconnected mitochondrial networks that can be seen in yeast mutants deficient in mitochondrial division¹⁷. Furthermore, in four out of four independent transgenic lines *P_{HS}drp-1(K40A)* expression caused the inappropriate survival of 2 or 3 of the 16 cells destined to die during the development of the anterior pharynx (Table 1)⁶. Strong loss-of-function mutations in the pro-apoptotic genes *egl-1*, *ced-4* or *ced-3* cause 11 or 12 of these cells to survive, whereas weak loss-of-function mutations in *egl-1*, *ced-4* or *ced-3* cause 2 or 3 of them to survive¹⁸. The cell-death defective (*Ced*) phenotype observed in lines expressing *P_{HS}drp-1(K40A)* is therefore comparable to the *Ced* phenotype caused by weak loss-of-function mutations in *egl-1*, *ced-4* or *ced-3*. A similar result was obtained when *drp-1(K40A)* was expressed under the control of the *egl-1* promoter (*P_{egl-1}drp-1(K40A)*) (Table 2). We conclude that programmed cell death can be inhibited by reducing DRP-1 activity and DRP-1-mediated fission of mitochondria.

We asked whether an increased rate of mitochondrial fission can induce programmed cell death. Overexpression of the wild-type

drp-1 gene, *drp-1(wt)*, in muscle cells of adult *C. elegans* enhances the fission of mitochondria, resulting in fragmentation of the organelles¹⁶. Overexpression of *drp-1(wt)* during embryogenesis using the heat-inducible promoter (*P_{HS}drp-1(wt)*) caused mitochondria to fragment in a similar manner (Fig. 2a, *P_{HS}drp-1(wt)*). In four out of five independent transgenic lines, the overexpression of *drp-1(wt)* resulted in a large increase in the number of refractile corpses detected in embryos (Table 3, Supplementary Fig. S3). To determine whether the appearance of refractile corpses was the result of ectopic cell death rather than a block in the engulfment of cells normally destined to die, we examined the fates of cells normally destined to survive. The neurosecretory motoneurons (NSMs), M3s and I3s are neurons in the anterior bulb of the pharynx that are destined to survive. We observed that about 20% of these cells died and became refractile upon *drp-1(wt)* expression ($n=37$ animals) (Fig. 2b, Supplementary Table S2). Furthermore, the appearance of corpses in embryos was suppressed by *egl-1(n1084 n3082)*, *ced-9(n1950gf)*, *ced-4(n1162)* or *ced-3(n717)*, confirming that they represented apoptotic cells (Table 3). We conclude that DRP-1-induced excessive mitochondrial fission causes ectopic programmed cell death (see Supplementary Tables S3 and S4 for additional data supporting this conclusion).

Like EGL-1-induced mitochondrial fragmentation, DRP-1-induced mitochondrial fragmentation was not blocked by

Table 3 Programmed cell death caused by expression of *drp-1(wt)*

Line	Average number of corpses	s.d.	Range	<i>n</i>
Control (1)	7.9	0.8	7–9	12
Control (2)	8.6	1.6	7–12	15
<i>P_{HS}drp-1(wt)</i> (1)	18.6	2.7	15–25	15
<i>P_{HS}drp-1(wt)</i> (2)	25.2	3.4	17–27	15
<i>P_{HS}drp-1(wt)</i> (3)	22.4	2.7	19–28	15
<i>P_{HS}drp-1(wt)</i> (4)	10.1	2.0	7–13	15
<i>P_{HS}drp-1(wt)</i> (5)	21.1	3.0	19–28	15
<i>egl-1(n1084 n3082)</i> ; <i>P_{HS}drp-1(wt)</i> (6)	0	0	0	7
<i>egl-1(n1084 n3082)</i> ; <i>P_{HS}drp-1(wt)</i> (7)	0	0	0	10
<i>ced-9(n1950gf)</i> ; <i>P_{HS}drp-1(wt)</i> (8)	1.1	1.0	0–3	10
<i>ced-9(n1950gf)</i> ; <i>P_{HS}drp-1(wt)</i> (9)	1.6	1.1	0–3	9
<i>+/+</i> ; <i>P_{HS}drp-1(wt)</i> (9*)	15.4	3.3	10–22	13
<i>ced-9(n1950gf)</i> ; <i>P_{HS}drp-1(wt)</i> (10)	2.7	1.1	0–4	10
<i>+/+</i> ; <i>P_{HS}drp-1(wt)</i> (10*)	17.9	3.9	10–22	13
<i>ced-4(n1162)</i> ; <i>P_{HS}drp-1(wt)</i> (11)	0	0	0	9
<i>ced-4(n1162)</i> ; <i>P_{HS}drp-1(wt)</i> (12)	0.1	0.3	0–1	11
<i>ced-4(n1162)</i> ; <i>P_{HS}drp-1(wt)</i> (13)	0.2	0.4	0–1	12
<i>ced-4(n1162)</i> ; <i>P_{HS}drp-1(wt)</i> (14)	0	0	0	15
<i>ced-3(n717)</i> ; <i>P_{HS}drp-1(wt)</i> (15)	0	0	0	20
<i>ced-3(n717)</i> ; <i>P_{HS}drp-1(wt)</i> (16)	0	0	0	26
<i>ced-3(n717)</i> ; <i>P_{HS}drp-1(wt)</i> (17)	0	0	0	6
<i>+/+</i> ; <i>P_{HS}drp-1(wt)</i> (17*)	19.6	4.1	12–24	15
<i>ced-3(n717)</i> ; <i>P_{HS}drp-1(wt)</i> (18)	0	0	0	30

Transgenes were activated and the numbers of refractile corpses were determined as described in Methods. s.d., standard deviation. Animals analysed were at the comma stage of embryogenesis. The complete genotypes of the animals were (from top to bottom): *unc-76(e911)*; *P_{HS}mitogfp* lines 1 and 2, *unc-76(e911)*; *P_{HS}mitogfp+P_{HS}drp-1(wt)* lines 1–5, *unc-76(e911)* *egl-1(n1084 n3082)*; *P_{HS}mitogfp+P_{HS}drp-1(wt)* lines 6 and 7, *ced-9(n1950gf)*; *unc-76(e911)*; *P_{HS}mitogfp+P_{HS}drp-1(wt)* lines 8, 9 and 10, *ced-4(n1162)*; *unc-76(e911)*; *P_{HS}mitogfp+P_{HS}drp-1(wt)* lines 11–14, *ced-3(n717)*; *unc-76(e911)*; *P_{HS}mitogfp+P_{HS}drp-1(wt)* lines 15–18. Lines marked with an asterisk were crossed out of the respective *ced* backgrounds to confirm that the transgenes were active in a wild-type background.

Table 2 Expression of *drp-1(K40A)* with *egl-1* promoter blocks cell death

Line	Average number of extra cells	s.d.	Range
Control (1)	0.3	0.5	0–1
Control (2)	0.3	0.5	0–1
Control (3)	0.1	0.4	0–1
Control (4)	0.1	0.4	0–1
Control (5)	0.1	0.4	0–1
Control (6)	0.2	0.4	0–1
Control (7)	0.1	0.3	0–1
<i>P_{egl-1}drp-1(K40A)</i> (1)	0.4	0.7	0–2
<i>P_{egl-1}drp-1(K40A)</i> (2)	2.3	1.1	1–4
<i>P_{egl-1}drp-1(K40A)</i> (3)	0.3	0.5	0–1
<i>P_{egl-1}drp-1(K40A)</i> (4)	3.1	1.7	1–7
<i>P_{egl-1}drp-1(K40A)</i> (5)	3.4	1.2	1–6
<i>P_{egl-1}drp-1(K40A)</i> (6)	2.8	1.2	0–5
<i>P_{egl-1}drp-1(K40A)</i> (7)	1.1	1.4	0–3

The number of extra cells in the anterior pharynx was determined as described in Methods. s.d., standard deviation. The complete genotypes of the animals were (from top to bottom): *unc-76(e911)*; *P_{HS}* control lines 1–7, *unc-76(e911)*; *P_{egl-1}drp-1(K40A)* lines 1–7.

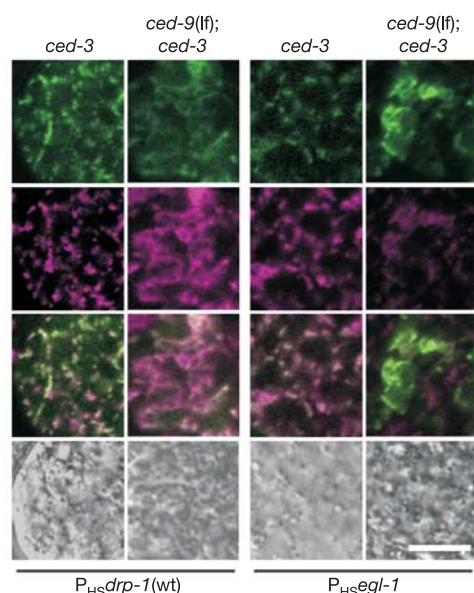


Figure 3 *ced-9(n2812lf)* blocks the ability of EGL-1 and DRP-1 to induce mitochondrial fragmentation. Representative confocal images are shown of mitoGFP, rhodamine, mitoGFP/rhodamine overlay and DIC (from top to bottom) of *ced-3(n717)* animals (first and third columns) or *ced-9(n2812lf); ced-3(n717)* animals (second and fourth columns) carrying a $P_{HS}drp-1(wt)$ and $P_{HS}mitogfp$ transgene (first and second columns) or a $P_{HS}egl-1$ and $P_{HS}mitogfp$ transgene (third and fourth columns). Transgenes were induced as described in Methods, and embryos were imaged at the comma to $1\frac{1}{2}$ -fold stage of embryonic development. Images represent single confocal image planes. Scale bar, 8 μ m.

ced-4(n1162) or *ced-3(n717)* (Fig. 2a). Furthermore, whereas *ced-9(n1950gf)* blocked DRP-1-induced mitochondrial fragmentation, *egl-1(n1084 n3082)* failed to do so. Thus, the overexpression of *drp-1(wt)* can bypass the requirement for *egl-1* function in mitochondrial fragmentation but not the ability of *ced-9(n1950gf)* to block mitochondrial fragmentation. *n1950gf* has been proposed to block programmed cell death by impairing the ability of the mitochondria-localized CED-9 protein to bind to EGL-1, thereby preventing the release of CED-4 in cells destined to die^{7,19}. The blockage of mitochondrial fragmentation by *ced-9(n1950gf)* could therefore be explained if CED-4 bound to CED-9 blocked mitochondrial fragmentation, if the binding of EGL-1 to CED-9 were essential for mitochondrial fragmentation, or if CED-9 were required for mitochondrial fragmentation in apoptotic cells and *n1950gf* impaired this activity of CED-9. Our observation that the overexpression of *drp-1(wt)* can bypass the requirement for *egl-1* function in mitochondrial fragmentation, but not the ability of *ced-9(n1950gf)* to block mitochondrial fragmentation, makes it unlikely that the latter is a result of impaired EGL-1 binding to CED-9. To distinguish between the two other possibilities, we analysed the effect of the *ced-9* loss-of-function mutation *n2812* on mitochondrial fragmentation induced by EGL-1 and DRP-1. Because of the loss of CED-9's cell-death protective function, *ced-9(n2812lf)* animals are not viable as a result of ectopic programmed cell death²⁰. Because *ced-3* acts downstream of *ced-9*, loss-of-function mutations of *ced-3*, such as *n717*, block *ced-9(n2812lf)*-induced ectopic programmed cell death and lethality²¹. As *ced-3(n717)* does not affect the ability of EGL-1 or DRP-1 to induce mitochondrial fragmentation (see above), we analysed the effect of *ced-9(n2812lf)* in the background of *ced-3(n717)*. We found that, like *ced-9(n1950gf)*, *ced-9(n2812lf)*

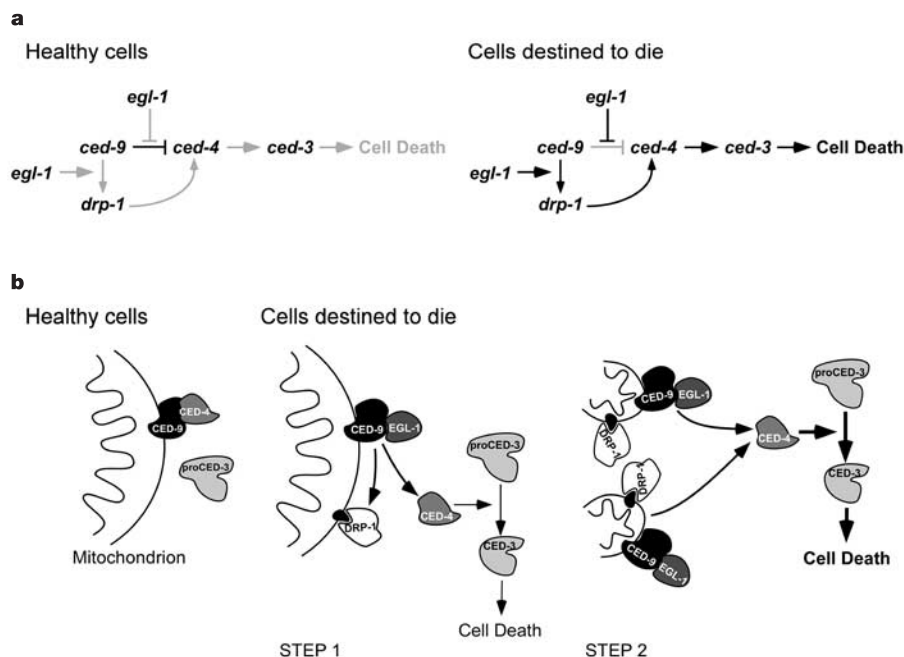


Figure 4 Genetic and molecular model of the activation of programmed cell death in *C. elegans*. **a**, A genetic pathway for the activation of programmed cell death during *C. elegans* development. In healthy cells, the anti-apoptotic function of *ced-9* prevents the activation of programmed cell death by negatively regulating the pro-apoptotic gene *ced-4*. In cells destined to die, *egl-1* blocks the anti-apoptotic function of *ced-9*, thereby activating *ced-4* and *ced-3*, which can result in programmed cell death. *egl-1* also activates the pro-apoptotic killing function of *ced-9*, thereby inducing *drp-1* activity, which contributes to programmed cell death by enhancing *ced-4* and *ced-3* activity. **b**, A simplified model for the molecular interactions occurring during the activation of

programmed cell death. In healthy cells, CED-9 blocks CED-4's ability to activate proCED-3. After EGL-1 binds to CED-9 in cells destined to die, CED-4 is released from mitochondria allowing CED-4 to activate proCED-3, which can result in programmed cell death (step 1). EGL-1 binding to CED-9 also activates CED-9's killing function, resulting in DRP-1 activation and DRP-1-mediated mitochondrial fragmentation. Mitochondrial fragmentation contributes to programmed cell death by potentiating CED-4 activity and proCED-3 activation (step 2). Alternatively, the pro-apoptotic activity of DRP-1 could be a function of DRP-1 that is independent of its function in mitochondrial fragmentation.

blocked the ability of EGL-1 or DRP-1 to induce mitochondrial fragmentation (Fig. 3). It has been shown that rather than localizing to mitochondria, CED-4 localizes to perinuclear membranes in *ced-9(n2812lf)* animals⁷. This result therefore rules out the possibility that it is the binding of CED-4 to CED-9(*n1950gf*) that blocks mitochondrial fragmentation in *ced-9(n1950gf)* animals. However, it indicates that *ced-9* is required for mitochondrial fragmentation in apoptotic cells and that *n1950gf* impairs this activity. Indeed, *ced-9* has been proposed to adopt a killing function in cells destined to die²². We propose that *ced-9*'s killing function in cells destined to die is equivalent to *ced-9*'s role in mitochondrial fragmentation. The pro-apoptotic Bcl-2 family member Bax induces mitochondrial fragmentation in mammalian cells and localizes to sites of mitochondrial fission^{23,24}. Bax might therefore be directly involved in the activation of mitochondrial fragmentation during apoptosis. In cells induced to die by the activation of *egl-1*, CED-9 might therefore be functionally homologous to Bax.

Studies on the *C. elegans* homologues of endonuclease G (EndoG) and apoptosis-inducing factor (AIF), namely CPS-6 (*cps*, for CED-3 protease suppressor) and WAH-1 (*wah*, for worm AIF homologue), respectively, have indicated a possible role for mitochondria late in the apoptotic process in *C. elegans* after CED-3 activation^{25,26}. Our data show that mitochondrial fragmentation is sufficient and, at least partly, required for the activation of programmed cell death during *C. elegans* development. Mitochondrial fragmentation in apoptotic cells is induced by the BH3-only protein EGL-1, dependent on the Bcl-2-like protein CED-9 and mediated by the dynamin-related protein DRP-1. In addition, mitochondrial fragmentation is presumably initiated before or simultaneously with the activation of the Apaf-1-like protein CED-4 and the caspase CED-3. These results provide the first evidence that mitochondria have an important function during an early stage of programmed cell death in *C. elegans*. How might mitochondrial fragmentation contribute to killing? DRP-1-induced programmed cell death is dependent on *egl-1*, *ced-4* and *ced-3*. Hence, the function of mitochondrial fragmentation in apoptotic cells might be to potentiate the activity of CED-4 proteins that have been released from mitochondria in an EGL-1-dependent manner, thereby enhancing CED-3 activation (Fig. 4). This observation raises the possibility that pro-apoptotic factors capable of increasing CED-4 activity might be released from mitochondria upon fragmentation. Our results assign an important function to mitochondria in programmed cell death in *C. elegans* and establish that mitochondrial fragmentation is an integral and conserved aspect of the cell-death activation machinery. □

Methods

General methods and strains

C. elegans strains were maintained at 20 °C on nematode growth medium (NGM) plates. N2 was the wild-type strain. Alleles used in this study are listed below and are described in ref. 27, except where noted otherwise: LGIII, *ced-4(n1162)*, *ced-9(n1950, n2812)*; LGIV, *ced-3(n717)*; LGV, *unc-76(e911)*, *egl-1(n1084 n3082)*¹³. *bcl-549* is a stable integration of the plasmids *P_{egl-1}mitogfp* and *p76-16B* (ref. 28). *bcl-51* is a stable integration of the plasmids *P_{egl-1}mitogfp* and *p76-16B*²⁸ (J. Hatzold and B.C., unpublished data). *P_{ced-3}dsRed* is expressed in five neurons in the anterior pharynx of embryos at the 3½–4-fold stage, the two NSMs, the two M3s, and the I3 (ref. 29), all of which are normally destined to survive.

Molecular biology

mitogfp was amplified by polymerase chain reaction from plasmid *myo-3::mitoGFP*¹⁶, cloned into pBluescript and used to replace *gfp* in pBC99 (ref. 8) (using *Xma*I and *Spe*I for subcloning) (*P_{egl-1}gfp*) or cloned into pPD49.78 and pPD49.83 (*C. elegans* vectors for heat-inducible expression) (using *Bam*HI and *Spe*I) to create *P_{egl-1}mitogfp* and *P_{HIS}mitogfp*, respectively. The pPD49.78-based and pPD49.83-based plasmids were injected together and are referred to as *P_{HIS}-drp-1(wt)*, and *drp-1(K40A)* were subcloned from plasmids *myo-3::drp-1* and *myo-3::drp-1K40A*¹⁶ into pBluescript (using *Bam*HI and *Spe*I), and cloned into pPD49.78 and pPD49.83 (using *Bam*HI and *Spe*I) and pBC99 (using *Xma*I and *Spe*I) to create *P_{HIS}drp-1*, *P_{HIS}drp-1(K40A)*, and *P_{egl-1}drp-1(K40A)*, respectively. *P_{HIS}egl-1* (pBC27 and pBC28) was described previously¹³.

Transgenic animals

Plasmids were injected (5 ng µl⁻¹) into *unc-76(e911)* mutants or into the wild type by

using the *unc-76(+)* rescuing plasmid *p76-16B*²⁸ or the plasmid *pRF4* (ref. 30) (which dominantly confers a Rol (for roller) phenotype) as co-injection marker, respectively (50–75 ng µl⁻¹). Transgenic non-Unc or Rol progeny were picked and used to establish lines. To create a stable line of animals carrying the *P_{egl-1}mitogfp* transgene, non-Unc animals were mutagenized with ethylmethanesulphonate. A stable integrant, *bcl-549*, was identified by screening for F₂ animals that transmitted the transgene to 100% of their progeny. For the expression of transgenes under the control of the heat-inducible promoter, heat shock was applied as follows: 10 gravid transgenic hermaphrodites were transferred to fresh plates, allowed to lay eggs for 1–2 h at 20 °C, heat-shocked for 45 min at 33 °C, permitted to recover for 1 h 15 min at 20 °C before being removed from the plates. Progeny laid during this 3-h period were analysed. Imaging was performed, and the number of refractile corpses in embryos was measured by confocal microscopy and DIC about 2 h after the application of the heat shock. Extra cells in the anterior pharynx were counted by DIC in transgenic L4 larvae that had been heat-shocked during embryogenesis.

Mitochondrial labelling and imaging

To reveal mitochondria in embryos we used a GFP with a sequence targeting the mitochondrial matrix (mitoGFP) and/or staining with potential-sensitive, mitochondria-selective rhodamine derivatives. For staining, L4 larvae or gravid adults were transferred to NGM plates containing 30 µM rhodamine B (hexyl ester, perchlorate; Molecular Probes) or 30 µM tetramethylrhodamine (ethyl ester, perchlorate; Molecular Probes). Progeny were mounted on slides with agar pads (2%) and M9 buffer. Mitochondrial morphology was analysed with a confocal microscope (Leica TCS SP2; Leica Lasertechnik, Bensheim, Germany; and Zeiss LSM 510; Carl Zeiss Microscopy, Jena, Germany) equipped with a 63 × objective and DIC setting. Image acquisition was performed at ambient temperature (about 22 °C). For time-lapse imaging, 512 × 512 pixel images at 0.5-µm focal increments were recorded every 5 min, simultaneously imaging mitoGFP, rhodamine and DIC over a period of 30–60 min. Images were processed with the Leica Confocal Software (LCS) or LCS Lite.

Received 29 September 2004; accepted 6 January 2005; doi:10.1038/nature03316.

- Horvitz, H. R. Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. *Cancer Res.* **59**, 1701s–1706s (1999).
- Daniel, N. N. & Korsmeyer, S. J. Cell death: critical control points. *Cell* **116**, 205–219 (2004).
- Newmeyer, D. D. & Ferguson-Miller, S. Mitochondria: releasing power for life and unleashing the machineries of death. *Cell* **112**, 481–490 (2003).
- Jiang, X. & Wang, X. Cytochrome C-mediated apoptosis. *Annu. Rev. Biochem.* **73**, 87–106 (2004).
- Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
- Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
- Chen, F. *et al.* Translocation of *C. elegans* CED-4 to nuclear membranes during programmed cell death. *Science* **287**, 1485–1489 (2000).
- Conradt, B. & Horvitz, H. R. The TRA-1A sex determination protein of *C. elegans* regulates sexually dimorphic cell deaths by repressing the *egl-1* cell death activator gene. *Cell* **98**, 317–327 (1999).
- Bossy-Wetzel, E., Barsoum, M. J., Godzik, A., Schwarzenbacher, R. & Lipton, S. A. Mitochondrial fission in apoptosis, neurodegeneration and aging. *Curr. Opin. Cell Biol.* **15**, 706–716 (2003).
- Karbowksi, M. & Youle, R. J. Dynamics of mitochondrial morphology in healthy cells and during apoptosis. *Cell Death Differ.* **10**, 870–880 (2003).
- Frank, S. *et al.* The role of dynamin-related protein 1, a mediator of mitochondrial fission, in apoptosis. *Dev. Cell* **1**, 515–525 (2001).
- Breckenridge, D. G., Stojanovic, M., Marcellus, R. C. & Shore, G. C. Caspase cleavage product of BAP31 induces mitochondrial fission through endoplasmic reticulum calcium signals, enhancing cytochrome c release to the cytosol. *J. Cell Biol.* **160**, 1115–1127 (2003).
- Conradt, B. & Horvitz, H. R. The *C. elegans* protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. *Cell* **93**, 519–529 (1998).
- Bloss, T. A., Witze, E. S. & Rothman, J. H. Suppression of CED-3-independent apoptosis by mitochondrial β NAC in *Caenorhabditis elegans*. *Nature* **424**, 1066–1071 (2003).
- van der Bliek, A. M. Functional diversity in the dynamin family. *Trends Cell Biol.* **9**, 96–102 (1999).
- Labrousse, A. M., Zappaterra, M. D., Rube, D. A. & van der Bliek, A. M. *C. elegans* dynamin-related protein DRP-1 controls severing of the mitochondrial outer membrane. *Mol. Cell* **4**, 815–826 (1999).
- Shaw, J. M. & Nunnari, J. Mitochondrial dynamics and division in budding yeast. *Trends Cell Biol.* **12**, 178–184 (2002).
- Reddien, P. W., Cameron, S. & Horvitz, H. R. Phagocytosis promotes programmed cell death in *C. elegans*. *Nature* **412**, 198–202 (2001).
- Parrish, J., Metters, H., Chen, L. & Xue, D. Demonstration of the *in vivo* interaction of key cell death regulators by structure-based design of second-site suppressors. *Proc. Natl Acad. Sci. USA* **97**, 11916–11921 (2000).
- Hengartner, M. O. & Horvitz, H. R. *C. elegans* cell survival gene *ced-9* encodes a functional homolog of the mammalian proto-oncogene bcl-2. *Cell* **76**, 665–676 (1994).
- Hengartner, M. O., Ellis, R. E. & Horvitz, H. R. *Caenorhabditis elegans* gene *ced-9* protects cells from programmed cell death. *Nature* **356**, 494–499 (1992).
- Hengartner, M. O. & Horvitz, H. R. Activation of *C. elegans* cell death protein CED-9 by an amino-acid substitution in a domain conserved in Bcl-2. *Nature* **369**, 318–320 (1994).
- Desagher, S. & Martinou, J. C. Mitochondria as the central control point of apoptosis. *Trends Cell Biol.* **10**, 369–377 (2000).
- Karbowksi, M. *et al.* Spatial and temporal association of Bax with mitochondrial fission sites, Drp1, and Mfn2 during apoptosis. *J. Cell Biol.* **159**, 931–938 (2002).
- Parrish, J. *et al.* Mitochondrial endonuclease G is important for apoptosis in *C. elegans*. *Nature* **412**, 90–94 (2001).
- Wang, X., Yang, C., Chai, J., Shi, Y. & Xue, D. Mechanisms of AIF-mediated apoptotic DNA degradation in *Caenorhabditis elegans*. *Science* **298**, 1587–1592 (2002).

27. Riddle, D. L. & Albert, P. S. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 739–768 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1997).
28. Bloom, L. & Horvitz, H. R. The *Caenorhabditis elegans* gene *unc-76* and its human homologs define a new gene family involved in axonal outgrowth and fasciculation. *Proc. Natl Acad. Sci. USA* **94**, 3414–3419 (1997).
29. Aspöck, G., Ruvkun, G. & Burglin, T. R. The *Caenorhabditis elegans* *ems* class homeobox gene *celh-2* is required for M3 pharynx motoneuron function. *Development* **130**, 3369–3378 (2003).
30. Mello, C. & Fire, A. DNA transformation. *Methods Cell Biol.* **48**, 451–482 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Lambie, H. Hutter and members of the Conradt laboratory for comments on the manuscript; H. Schnabel and J. Hatzold for discussion; J. Hatzold for generating the integration *bcl51*; A. van der Bliek for providing the plasmids *myo-3::mitogfp*, *myo-3::drp-1* and *myo-3::drp-1(K40A)*; S. Hell, S. Jakobs, G. Tavasani, A. Vollmar and J. Chalcroft for their support with microscopy; and W. Neupert for his support throughout this study. This research was supported by funding from the Deutsche Forschungsgemeinschaft and the Friedrich-Baur-Stiftung to B.W. and by funding from the Max Planck Society, the European Molecular Biology Organization (EMBO Young Investigator Award) and the Howard Hughes Medical Institute (HHMI award to Dartmouth Medical School under the Biomedical Research Support Program for Medical Schools) to B.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.C. (barbara.conradt@dartmouth.edu).

Rejuvenation of aged progenitor cells by exposure to a young systemic environment

Irina M. Conboy^{1,†}, Michael J. Conboy^{1,*}, Amy J. Wagers^{2,†}, Eric R. Girma¹, Irving L. Weissman² & Thomas A. Rando^{1,3}

¹Department of Neurology and Neurological Sciences and ²Department of Pathology, Stanford University School of Medicine, Stanford, California 94305, USA

³GRECC and Neurology Service, VA Palo Alto Health Care System, Palo Alto, California 94304, USA

* These authors contributed equally to this work

† Present addresses: Department of Bioengineering, University of California–Berkeley, Berkeley, California 94720, USA (I.M.C.); Section on Developmental and Stem Cell Biology, Joslin Diabetes Center, Boston, Massachusetts 02215, USA (A.J.W.)

The decline of tissue regenerative potential is a hallmark of ageing and may be due to age-related changes in tissue-specific stem cells^{1–5}. A decline in skeletal muscle stem cell (satellite cell) activity due to a loss of Notch signalling results in impaired regeneration of aged muscle^{1,6}. The decline in hepatic progenitor cell proliferation owing to the formation of a complex involving cEBP- α and the chromatin remodelling factor brahma (Brm) inhibits the regenerative capacity of aged liver⁷. To examine the influence of systemic factors on aged progenitor cells from these tissues, we established parabiotic pairings (that is, a shared circulatory system) between young and old mice (heterochronic parabioses), exposing old mice to factors present in young serum. Notably, heterochronic parabiosis restored the activation of Notch signalling as well as the proliferation and regenerative capacity of aged satellite cells. The exposure of satellite cells from old mice to young serum enhanced the expression of the Notch ligand (Delta), increased Notch activation, and enhanced proliferation *in vitro*. Furthermore, heterochronic parabiosis increased aged hepatocyte proliferation and restored the cEBP- α complex to levels seen in young animals. These results suggest that the age-related decline of progenitor cell activity can be modulated by systemic factors that change with age.

Ageing of metazoans can be generally characterized as a progressive decline of tissue and organ function, accompanied by increased oxidative damage⁸, mitochondrial dysfunction⁹, endocrine imbalance¹⁰ and genome instability¹¹. Tissue regenerative capacity also declines with age, and in tissues such as muscle, blood, liver and brain this decline has been attributed to a diminished responsiveness of tissue-specific stem and progenitor cells^{1,3–5}. Our previous work demonstrated that signalling through the Notch pathway is essential for the activation, proliferation and myogenic lineage progression of satellite cells necessary for muscle repair⁶, and that the decline in the regenerative potential of muscle with age is due to the failure of this pathway to be activated¹. However, the activation of aged satellite cells and the regenerative potential of aged muscle can be restored by forced activation of Notch, demonstrating that the intrinsic regenerative capacity of aged satellite cells remains intact¹. Aged muscle successfully regenerates when grafted into muscle in a young host, but young muscle displays impaired regeneration when grafted into an aged host^{12,13}. We hypothesized that there are systemic factors that support the robust regeneration of tissues in young animals and/or inhibit regeneration in old animals, and that these factors act to modulate the key molecular pathways that control the regenerative properties of progenitor cells. The implication of this hypothesis is that old tissues might be made to regenerate as well as young tissues if, by means of systemic influences, the molecular pathways could be ‘rejuvenated’ from an old state to a young state.

To test this hypothesis we set up an experimental system in which—in contrast to transplantation—regenerating tissues in aged animals could be exposed only to the circulating factors of young animals, and vice versa. We established parabiotic pairings between young and old mice (heterochronic parabioses)¹⁴, with parabiotic pairings between two young mice or two old mice (isochronic parabioses) as controls. In parabiosis, animals develop vascular anastomoses and thus a single, shared circulatory system. Previous work examining the effects of heterochronic parabiosis showed that such pairings may alter tissue function^{15–17}, but the effects on progenitor cell activity or tissue regeneration have not been examined.

We examined the efficacy of muscle regeneration in young (2–3 months) and aged (19–26 months) mice in heterochronic and isochronic pairings. Young partners were either transgenic for green fluorescent protein (GFP), expressed from the constitutive β -actin promoter, or expressed a distinct CD45 allele (Ly5.2/CD45.1), allowing confirmation of blood chimaerism (Supplementary Fig. S1)¹⁸. The use of GFP-transgenic mice as one member of a pair also allowed us to distinguish the cells from each animal participating in tissue regeneration. After 5 weeks of parabiosis the hindlimb muscles of each mouse were injured and the mice were given 5-bromodeoxyuridine (BrdU) injections⁶. After muscle injury, satellite cells activate and give rise to proliferative myoblasts, which ultimately fuse to form nascent myofibres (myotubes) that maintain centrally located nuclei and express embryonic myosin heavy chain (eMHC), a specific marker of regenerating myotubes in adult animals. Five days after injury, muscles in young mice in both isochronic and heterochronic parabioses had regenerated robustly, as demonstrated by the appearance of centrally nucleated, eMHC-expressing myotubes (Fig. 1a) formed from proliferating progenitor cells (Supplementary Fig. S2). In contrast, injured muscle from old isochronic parabionts regenerated poorly, typical of aged animals¹, with a failure of myotube formation, prominent fibrosis at the site of injury, and evidence of proliferating cells predominantly in the interstitial spaces (Supplementary Fig. S2). Notably, parabiosis with young mice significantly enhanced the regeneration of muscle in old partners (Fig. 1a, b; see also Supplementary Fig. S2). The appearance of nascent myotubes in these old mice was similar to that seen in young mice.

Importantly, the enhanced regeneration of aged muscle was due

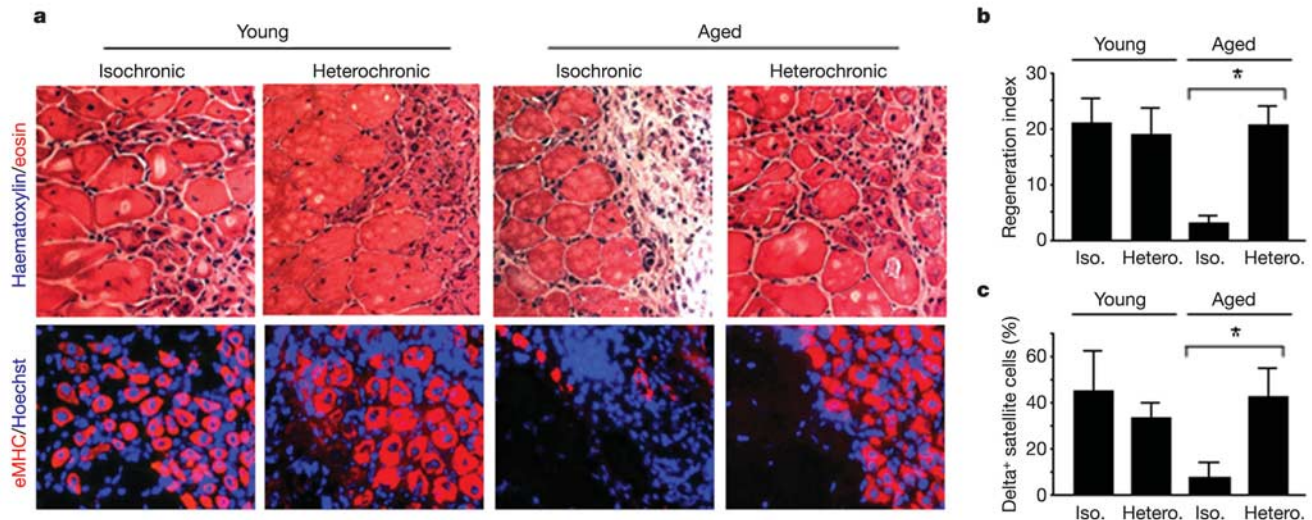


Figure 1 Heterochronic parabiosis restores muscle regeneration and muscle stem cell activation in aged animals. **a**, Five days after injury, muscles from parabiotic mice were analysed for indices of regeneration and nascent myotube formation by conventional histological analysis (top panels: haematoxylin and eosin staining) and by immunostaining for eMHC (bottom panels: eMHC (red); Hoechst dye (blue) labels all nuclei). Representative fields are shown. **b**, Quantification of the effectiveness of the regenerative responses from

experiments represented in **a** (see Methods). Data are from three or more pairs for each parabiotic condition (asterisk, $P < 0.005$). **c**, After 5 weeks of parabiosis, bulk muscle fibres were isolated and cultured *ex vivo* for 1 day. Satellite cells were prepared and analysed by flow cytometry for Delta and CD34 expression to determine the percentage of activated satellite cells (that is, those CD34⁺ cells with high Delta expression⁶) ($n = 4-6$ for each condition; asterisk, $P < 0.005$). Error bars indicate standard deviation.

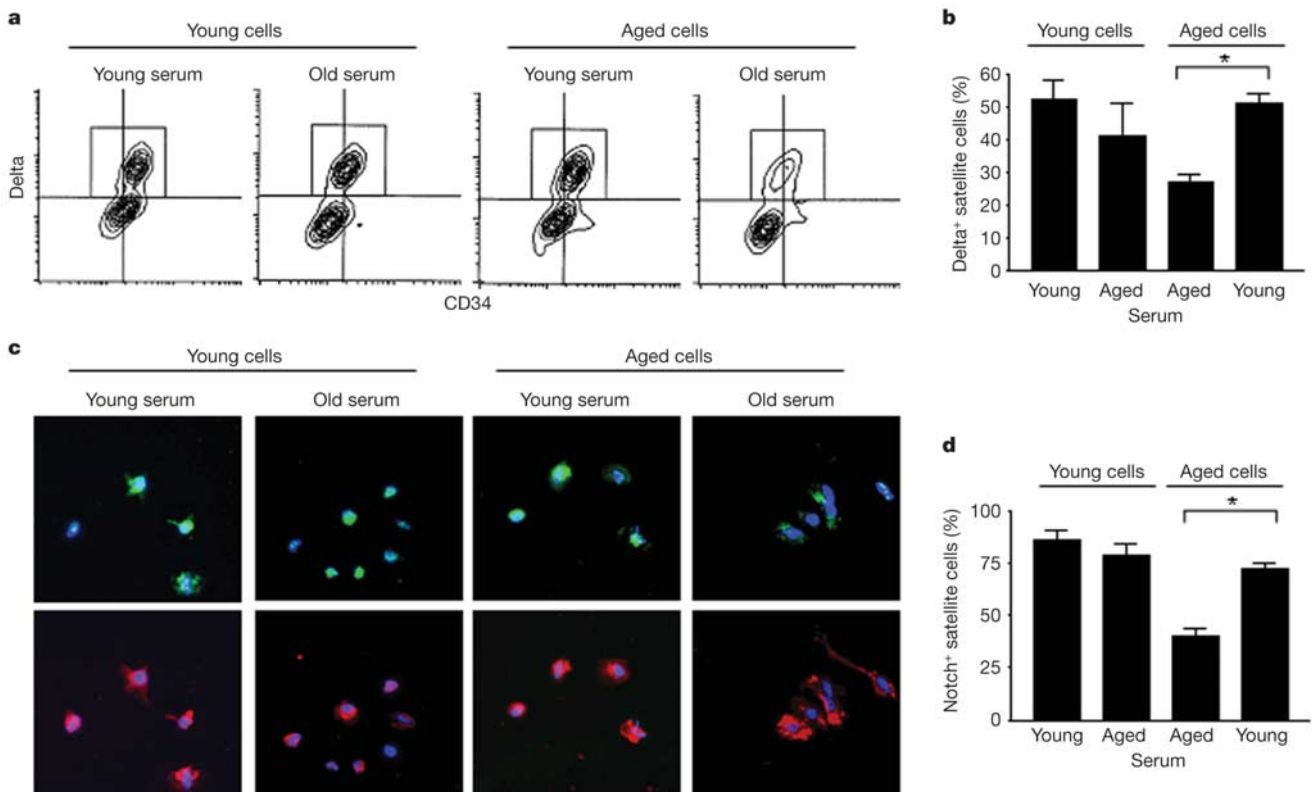


Figure 2 Young serum rejuvenates activation of aged satellite cells. **a**, Myofibre-associated satellite cells were isolated from young or old mice, cultured overnight in media containing either young or old serum, and then analysed by flow cytometry for the expression of Delta and CD34, as described¹. The data clearly discriminate between Delta-positive (above the horizontal line) and Delta-negative cells. Satellite cells are CD34-positive (to the right of the vertical line) upon isolation, but CD34 expression declines in the population with time in culture, as described¹. After 24 h *ex vivo*, some cells have already become CD34-negative, an effect accentuated in old serum.

b, Quantification of satellite cell activation from experiments represented in **a**. $n = 4$; asterisk, $P < 0.005$. **c**, Fibre-associated cells from young or old mice were plated in media containing young or old serum for 1 day before fixation. Cells were immunostained for active Notch-1 (green) and M-cadherin (a satellite cell marker²⁶; red). Hoechst dye (blue) labels all nuclei. **d**, Cells from experiments such as in **c** were quantified to determine the number of satellite cells (M-cadherin-positive cells) in which activated Notch was detected ($n \geq 3$; asterisk, $P < 0.005$). Error bars indicate standard deviation.

almost exclusively to the activation of resident, aged progenitor cells, not to the engraftment of circulating progenitor cells from the young partner. Less than 0.1% of regenerated myotubes in aged muscle were GFP-positive (Supplementary Fig. S4a–c), consistent with comparable studies testing for the contribution of circulating cells to muscle regeneration^{18–20}. The GFP transgene has been shown not only to be expressed robustly in mature myofibres of the transgenic strain, but cells isolated from this strain express GFP after incorporation into a GFP-negative recipient^{18,19,21}. Therefore, this is clearly a sensitive and accurate measure of engraftment. Nearly all GFP-positive cells present at the site of the injury were located in the interstitial spaces (Supplementary Fig. S4b), and these cells were probably infiltrating leukocytes, as has been described previously^{18,21}. These results indicate that the impaired regenerative potential of aged satellite cells can be improved by a modification of the systemic environment, by means of an increase of positive factors in young mouse serum, a decrease or dilution of inhibitory factors present in old mouse serum, or both.

The loss of muscle regeneration with age is due at least in part to an age-related impairment in the upregulation of the Notch ligand Delta after muscle injury¹. Therefore, we tested whether heterochronic parabiosis restored Delta upregulation in aged satellite cells and thus enhanced their activation and proliferation. Using myofibre explantation to assess satellite cell activation^{6,22}, we analysed satellite cells for the expression of Delta. In young isochronic and heterochronic parabionts there was a marked upregulation of Delta in satellite cells, whereas Delta induction was lacking in the old, isochronic parabionts (Fig. 1c), typical of the response of aged muscle¹. Notably, satellite cells from the aged partners of heterochronic parabionts showed a marked upregulation of Delta, comparable to that found in their young partners (Fig. 1c) and in young mice not subjected to parabiotic pairings¹. There was a slight inhibition of Delta upregulation in satellite cells from young mice in heterochronic parabioses compared with young isochronic parabionts (Fig. 1c). Thus, heterochronic parabiosis not only enhances the proliferative response of the resident aged progenitor cells, it also restores the key molecular signalling in these cells that is necessary for muscle regeneration¹.

To determine whether the rejuvenating effects of heterochronic parabiosis could be replicated in an *in vitro* system, satellite cells, cultured either alone or in association with myofibre explants⁶, were prepared from young and old mice and were cultured in the presence of young or old mouse serum, thus recapitulating the humoral aspects of heterochronic parabioses. Compared with young satellite cells cultured in either young or old serum, there was much less upregulation of Delta in old satellite cells cultured with old mouse serum (Fig. 2a, b); however, young mouse serum restored the upregulation of Delta (Fig. 2a, b) and the activation of Notch (Figs 2c, d) in aged satellite cells. There was an inhibitory effect of the old mouse serum on young satellite cells in terms of the upregulation of Delta (Fig. 2b). Serum from young mice also significantly enhanced the proliferation of myofibre-associated satellite cells in the old cultures compared with serum from old mice (Supplementary Fig. S3). There was reduced proliferation of both young and old satellite cells cultured in young mouse serum when Notch signalling was inhibited (Supplementary Fig. S3). Therefore, enhanced activation and proliferation of aged myogenic progenitor cells by serum from young mice is Notch dependent. Together, these *in vivo* and *in vitro* data demonstrate that components in young serum alone are capable of reversing the molecular and cellular aspects of age-related decline in muscle stem cell activation. In addition, it also seems that factors in old serum negatively affect the processes required for satellite cell activation and muscle repair.

Encouraged by these findings, we then examined liver from aged mice subjected to heterochronic parabiosis to test for evidence of a more general capability to rejuvenate aged, resident progenitor cells.

In liver, there are multiple types of progenitor cells that participate in regeneration and homeostasis, depending on the physiological or pathological condition²³. Here, we studied proliferating hepatocytes involved in normal tissue turnover, because there is a well documented decline in hepatocyte proliferation with age^{4,7,24}. Heterochronic and isochronic parabionts were examined for hepatocyte proliferation by two independent criteria: the incorporation of BrdU or expression of the proliferation marker Ki67 in albumin-positive cells. In young isochronic parabionts (Fig. 3a), the levels of basal hepatocyte proliferation were two- to threefold greater than in non-parabiosed controls (see Supplementary Fig. S5a), consistent with previous reports of effects of parabiosis on hepatocytes^{15,16}. Also, clusters of BrdU-positive proliferating cells that were not hepatocytes were observed in aged livers (Supplementary Fig. S5b). Proliferation of albumin-positive cells in old isochronic parabionts was less than that observed in young isochronic parabionts (Fig. 3a), consistent with previous reports^{4,7} and our observation (Supplementary Fig. S5a) that there is an age-related decline in the basal rate of hepatocyte proliferation in non-parabiosed animals. However, parabiosis to a young partner significantly increased

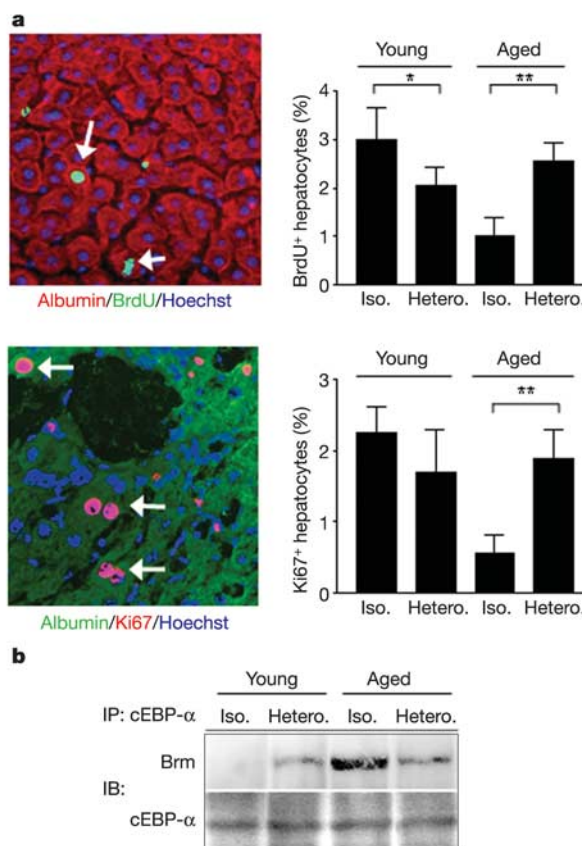


Figure 3 Heterochronic parabiosis enhances proliferation of aged liver progenitor cells and restores molecular determinants of young liver regeneration. **a**, After 5 weeks of parabiotic pairing, mice were given BrdU injections 5, 3 and 1 day before they were killed. Liver sections were analysed for hepatocyte proliferation based on co-staining with albumin and either BrdU or Ki67 (arrows point to nuclei of double-labelled cells). Hoechst dye (blue) labels all nuclei. Quantification of hepatocyte proliferation (see Methods) is shown to the right of each panel (asterisk, $P < 0.05$; double asterisk, $P < 0.005$; $n = 3–5$). **b**, Extracts from liver sections of parabionts were immunoprecipitated with an anti-cEBP- α antibody and the levels of Brm and cEBP- α were assessed by western blot analysis. Similar amounts of cEBP- α were detected after immunoprecipitation and similar levels of total Brm were present in all liver extracts (not shown), confirming that the formation of the complex, not the levels of the individual components, was affected by age and modified by heterochronic parabiosis. Similar results were obtained in four independent experiments. Error bars indicate standard deviation.

hepatocyte proliferation in aged mice (Fig. 3a). As in muscle, a reproducible reduction of progenitor cell proliferation was detected in livers of young mice after parabiotic pairing with old mice (Fig. 3a). Also as in muscle, the enhancement of hepatocyte proliferation in aged mice was due to resident cells and not the engraftment of circulating cells from the young partner. Consistent with previous observations¹⁸, less than 0.1% of hepatocytes expressed GFP in the non-transgenic, parabiotic partners.

The age-related decline in hepatocyte proliferation is due to the formation of an age-specific complex of cell cycle regulators associated with cEBP- α that inhibits E2F-driven gene expression⁷. One of the proteins detected in this complex in old, but not young, liver is the chromatin remodelling factor Brm. Levels of the cEBP- α -Brm complex increase in the livers of old rodents, leading to a decline of hepatocyte proliferation⁷. We tested whether the enhancement of aged hepatocyte proliferation correlated with reduced levels of cEBP- α -Brm complex formation in the aged mice subjected to heterochronic parabiosis. The cEBP- α -Brm protein complex was detected in liver extracts from old isochronic parabionts but not from young isochronic parabionts (Fig. 3b), similar to previous reports in non-parabiosed rodents⁷. Notably, the formation of the cEBP- α -Brm complex was diminished in liver extracts from old heterochronic parabionts (Fig. 3b). Thus, the molecular determinant had reverted to the 'youthful' phenotype, consistent with the enhanced proliferative activity of hepatocytes in these mice. The complex was present at elevated levels in young heterochronic parabionts compared with young controls, also consistent with the modest inhibition of hepatocyte proliferation in young heterochronic parabionts. These data show that the young systemic environment restores a younger profile of molecular signalling to the aged progenitor cells in the liver and also enhances their proliferation.

Our experiments suggest that there are systemic factors that can modulate the molecular signalling pathways critical to the activation of tissue-specific progenitor cells, and that the systemic environment of a young animal is one that promotes successful regeneration, whereas that of an older animal either fails to promote or actively inhibits successful tissue regeneration. It will be of great interest to identify the factors that have such a critical influence on tissue-specific progenitor cells. Our studies also demonstrate that the decline of tissue regenerative potential with age can be reversed through the modulation of systemic factors, suggesting that tissue-specific stem and progenitor cells retain much of their intrinsic proliferative potential even when old, but that age-related changes in the systemic environment and niche in which progenitor cells reside preclude full activation of these cells for productive tissue regeneration. □

Methods

Animal strains and parabiosis

Young C57BL/Ka-Ly5.2 and β -actin-eGFP transgenic mice were bred and maintained by the Weissman Laboratory at the Stanford University Research Animal Facility. GFP-transgenic mice were generated as described²⁵, and were backcrossed for at least ten generations to C57BL/Ka-Thy1.1 mice. Old C57BL/6 mice were obtained from the National Institutes of Aging or aged on site; young C57BL/6 mice were obtained from Jackson Laboratories. Animals were housed and handled in accordance with the guidelines of Veterinary Medical Unit of the VA Palo Alto Health Care System and the Administrative Panel on Laboratory Animal Care of Stanford University. Parabiosis was performed as published¹⁸; the viability of the pairs was similar to published data¹⁴.

Reagents

Antibodies to Delta and cEBP- α were purchased from Santa Cruz Biotechnology, to Ki67 from DakoCytomation, to CD34 and Brm from Beckton-Dickinson, to albumin and laminin from Sigma, and to active Notch-1 from Novus Biologicals/ABCAM. The antibody to eMHC was provided by G. Pavlath. The anti-BrdU staining kit was from Beckton-Dickinson and the BrdU-labelling reagent was from Zymed and was used at a dose of 50 $\mu\text{g g}^{-1}$. Soluble Delta-4 and Jagged-1 were from R&D systems. Fluorophore-conjugated secondary antibodies for flow cytometry and immunofluorescence were obtained from Caltag and used as previously described⁶.

Muscle injury

In order to analyse satellite cell activation and muscle regeneration *in vivo*, a small focal injury was made by an application of dry ice for 5 s directly to the tibialis anterior muscle¹. This generates a reproducible 'wedge' injury in the muscle with a discrete border between uninjured and injured muscle, and this border remains clear and distinct during the regeneration of the injured tissue.

Immunofluorescence and histological analysis

Muscles were dissected, embedded, cryo-sectioned and immunostained as previously described⁶. Liver was similarly prepared for cryo-sectioning and immunohistological analysis. Livers that showed signs of gross pathological abnormalities were excluded before analysis.

The effectiveness of the muscle regeneration was assessed by counting the number of eMHC-positive fibres at the injury site, and normalizing this to the total number of nuclei in the field of view in 7–12 randomly selected sections. This is presented as the 'regeneration index' and reflects both the positive (myotube formation) and negative (persistent inflammatory cell infiltration and fibroblast proliferation) aspects of regeneration that are present in young and aged mice.

To quantify hepatocyte proliferation, the total number of albumin-positive cells that were also either BrdU-positive or Ki67-positive were counted and expressed as a percentage of total albumin-positive cells in random fields. Ki67-positive or BrdU-positive cells that were negative for albumin were not included (see Supplementary Fig. S5b).

Satellite cell activation and isolation

In order to induce activation of satellite cells reproducibly and robustly, hindlimb muscles (tibialis anterior, gastrocnemius, biceps femoris and quadriceps) were dissected, digested into bulk fibres, and cultured overnight *ex vivo*¹. Satellite cells were purified from bulk fibres as described¹. The resulting preparation contained mononucleated cells that were >95% CD34⁺/M-cadherin⁺ satellite cells, which could be distinguished from remaining debris using flow cytometry by the characteristic forward scatter and side scatter properties of cells¹.

Satellite cell culture *in vitro*

For *in vitro* studies of satellite cell activation in the presence of mouse serum, bulk fibres were cultured at a ratio of one volume of loosely settled fibres to no less than 20 volumes of OptiMem (plus penicillin/streptomycin (Gibco/Invitrogen)), supplemented with 5% serum from mice of the indicated age¹. No additional matrix or growth factors were added to the fibres.

For detection of activated Notch, satellite cells were digested from muscle fibres as described¹ and plated onto ECM-coated (ECM gel, Sigma; diluted 1:1,000 in ice-cold PBS, applied at 1 $\mu\text{g cm}^{-2}$) chamber slides (Permanox or glass 8-well, Nalge-Nunc), and cultured at low density in OptiMEM with 5% mouse serum as above until fixation in 4% paraformaldehyde/PBS for 5 min for analysis.

For satellite cell proliferation, fibres were cultured as above with one-half of the media replaced every day. After addition of BrdU to the medium for 2 h, the percentages of BrdU-positive cells were determined as the total number of BrdU-positive nuclei per the total number of nuclei in fibres with associated BrdU-positive cells.

Flow cytometry

Cells were analysed either live or after fixation with 4% paraformaldehyde. Cells were stained with primary antibodies or with isotype-matched control antibodies (1 μg in 100 μl), followed by incubation with fluorochrome-labelled secondary antibodies (1:200–1:500 dilutions). Incubations were performed for 30 min on ice (live cells) or for 1 h at room temperature (fixed cells). Cells were analysed by flow cytometry (FACScan, Beckton Dickinson) using compensation for green and red fluorescence and doublet discrimination parameters.

Immunoprecipitation and western blotting

Immunoprecipitation of cEBP- α was from collected liver sections as described⁷. Western analysis was performed as previously described¹.

Statistical analysis

Data are presented as means and standard deviations. Comparisons between groups were done using Student's *t*-test assuming two-tailed distribution and unequal variances.

Received 14 June; accepted 9 December 2004; doi:10.1038/nature03260.

- Conboy, I. M., Conboy, M. J., Smythe, G. M. & Rando, T. A. Notch-mediated restoration of regenerative potential to aged muscle. *Science* **302**, 1575–1577 (2003).
- Morrison, S. J., Wandycz, A. M., Akashi, K., Globerson, A. & Weissman, I. L. The aging of hematopoietic stem cells. *Nature Med.* **2**, 1011–1016 (1996).
- Fuller, J. Hematopoietic stem cells and aging. *Sci. Aging Knowledge Environ.*, e11 (2002).
- Sigal, S. H., Brill, S., Fiorino, A. S. & Reid, L. M. The liver as a stem cell and lineage system. *Am. J. Physiol.* **263**, G139–G148 (1992).
- Kuhn, H. G., Dickinson-Anson, H. & Gage, F. H. Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *J. Neurosci.* **16**, 2027–2033 (1996).
- Conboy, I. M. & Rando, T. A. The regulation of Notch signaling controls satellite cell activation and cell fate determination in postnatal myogenesis. *Dev. Cell* **3**, 397–409 (2002).
- Iakova, P., Awad, S. S. & Timchenko, N. A. Aging reduces proliferative capacities of liver by switching pathways of C/EBP α growth arrest. *Cell* **113**, 495–506 (2003).
- Golden, T. R., Hinerfeld, D. A. & Melov, S. Oxidative stress and aging: beyond correlation. *Aging Cell* **1**, 117–123 (2002).
- Hekimi, S. & Guarente, L. Genetics and the specificity of the aging process. *Science* **299**, 1351–1354 (2003).

10. Tatar, M., Bartke, A. & Antebi, A. The endocrine regulation of aging by insulin-like signals. *Science* **299**, 1346–1351 (2003).
11. Hasty, P., Campisi, J., Hoeijmakers, J., van Steeg, H. & Vijg, J. Aging and genome maintenance: lessons from the mouse? *Science* **299**, 1355–1359 (2003).
12. Carlson, B. M. & Faulkner, J. A. Muscle transplantation between young and old rats: age of host determines recovery. *Am. J. Physiol.* **256**, C1262–C1266 (1989).
13. Carlson, B. M., Dedkov, E. I., Borisov, A. B. & Faulkner, J. A. Skeletal muscle regeneration in very old rats. *J. Gerontol. A* **56**, B224–B233 (2001).
14. McCay, C. M., Pope, F., Lunsford, W., Sperling, G. & Sambhavaphol, P. Parabiosis between old and young rats. *Gerontologia* **1**, 7–17 (1957).
15. Finerty, J. Parabiosis in physiological studies. *Physiol. Rev.* **32**, 277–302 (1952).
16. Tauchi, H. & Hasegawa, K. Change of the hepatic cells in parabiosis between old and young rats. *Mech. Ageing Dev.* **6**, 333–339 (1977).
17. Sidorenko, A. V., Gubrii, I. B., Andrianova, L. E., Macsijuk, T. V. & Butenko, G. M. Functional rearrangement of lymphohemopoietic system in heterochronically parabiosed mice. *Mech. Ageing Dev.* **36**, 41–56 (1986).
18. Wagers, A. J., Sherwood, R. L., Christensen, J. L. & Weissman, I. L. Little evidence for developmental plasticity of adult hematopoietic stem cells. *Science* **297**, 2256–2259 (2002).
19. Sherwood, R. L. *et al.* Isolation of adult mouse myogenic progenitors: Functional heterogeneity of cells within and engrafting skeletal muscle. *Cell* **119**, 543–554 (2004).
20. Partridge, T. A. Cells that participate in regeneration of skeletal muscle. *Gene Ther.* **9**, 752–753 (2002).
21. Sherwood, R. L., Christensen, J. L., Weissman, I. L. & Wagers, A. J. Determinants of skeletal muscle contributions from circulating cells, bone marrow cells, and hematopoietic stem cells. *Stem Cells* **22**, 1292–1304 (2004).
22. Rosenblatt, J. D., Lunt, A. I., Parry, D. J. & Partridge, T. A. Culturing satellite cells from living single muscle fiber explants. *In Vitro Cell. Dev. Biol. Anim.* **31**, 773–779 (1995).
23. Sell, S. Heterogeneity and plasticity of hepatocyte lineage cells. *Hepatology* **33**, 738–750 (2001).
24. Timchenko, N. A. *et al.* Regenerating livers of old rats contain high levels of C/EBP α that correlate with altered expression of cell cycle associated proteins. *Nucleic Acids Res.* **26**, 3293–3299 (1998).
25. Wright, D. E. *et al.* Cyclophosphamide/granulocyte colony-stimulating factor causes selective mobilization of bone marrow hematopoietic stem cells into the blood after M phase of the cell cycle. *Blood* **97**, 2278–2285 (2001).
26. Beauchamp, J. R. *et al.* Expression of CD34 and Myf5 defines the majority of quiescent adult skeletal muscle satellite cells. *J. Cell Biol.* **151**, 1221–1234 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. Chan and K. Robinson for technical help, and T. Wyss-Coray, T. Palmer, B. Omary and M. Buckwalter for discussions. The work was supported by grants from the Burroughs Wellcome Fund Career Award to A.J.W., and from the NIH, the American Federation for Aging Research (Paul Beeson Faculty Scholar in Aging) and the Department of Veterans Affairs (Merit Review) to T.A.R.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to T.A.R. (rand@stanford.edu).

Mitf cooperates with Rb1 and activates *p21^{Cip1}* expression to regulate cell cycle progression

Suzanne Carreira¹, Jane Goodall¹, Isil Aksan^{1*}, S. Anna La Rocca^{1*}, Marie-Dominique Galibert^{1*}, Laurence Denat², Lionel Larue² & Colin R. Goding¹

¹Signalling and Development Laboratory, Marie Curie Research Institute, The Chart, Oxted, Surrey RH8 0TL, UK

²UMR146 CNRS, Institut Curie, Bât. 110, Centre Universitaire 91405, Orsay cedex, France

* Present addresses: Yeditepe University, Faculty of Engineering and Architecture, Department of Genetics and Bioengineering, 26 Agustos Yerlesimi, Kayisdagi, 34755 Istanbul, Turkey (I.A.); Department of Immunology and Pathology, Institute of Animal Health, Pirbright, Surrey GU24 0NF, UK (S.A.L.R.); Génétique et Développement, CNRS UMR6061, Faculté de Médecine, Université Rennes, 1,2 avenue Léon Bernard, 35043 Rennes cedex, France (M.-D.G.)

The controls that enable melanoblasts and melanoma cells to proliferate are likely to be related, but so far no key regulator of cell cycle progression specific to the melanocyte lineage has been identified. The microphthalmia-associated transcription factor Mitf has a crucial but poorly defined role in melanoblast and

melanocyte survival and in differentiation¹. Here we show that Mitf can act as a novel anti-proliferative transcription factor able to induce a G1 cell-cycle arrest that is dependent on Mitf-mediated activation of the *p21^{Cip1}* (CDKN1A) cyclin-dependent kinase inhibitor gene. Moreover, cooperation between Mitf and the retinoblastoma protein Rb1 potentiates the ability of Mitf to activate transcription. The results indicate that Mitf-mediated activation of *p21^{Cip1}* expression and consequent hypophosphorylation of Rb1 will contribute to cell cycle exit and activation of the differentiation programme. The mutation of genes associated with melanoma, such as *INK4a* or *BRAF* that would affect either Mitf cooperation with Rb1 or Mitf stability respectively, would impair Mitf-mediated cell cycle control.

Treatment of melanocytes with forskolin leads to an accumulation of cells in G1 (Fig. 1a). Because activation of the cyclic AMP pathway induced by forskolin and melanocyte-stimulating hormone results in an elevated expression of Mitf² (Fig. 1a), we investigated the possibility that Mitf might regulate cell cycle progression. Preliminary experiments (not shown) established that even low levels of ectopic expression of Mitf in melanocyte cell lines that express endogenous Mitf were detrimental to growth even though these cells were *INK4a*-negative. We therefore used Mitf-negative NIH 3T3 cells to establish lines stably expressing Mitf (Mitf-M(+)) fused to the 4-hydroxytamoxifen (4-OHT)-responsive ligand-binding domain of the oestrogen receptor (ER)³. Immunofluorescence (Fig. 1b) revealed that most ER-Mitf was cytoplasmic but accumulated in the nucleus after treatment with 4-OHT. The addition of 4-OHT to cells expressing ER-Mitf led to an altered morphology reminiscent of melanocytes (Fig. 1c), although no induction of melanocyte markers such as *Tyrosinase* or *Tyrp-1* was observed (data not shown), and there was a reduced growth rate (Fig. 1d) compared with control NIH 3T3 cells expressing the ER ligand-binding domain only. Even in the absence of 4-OHT the ER-Mitf cells grew more slowly than the ER-only control cells in the presence or absence of 4-OHT, most probably because a small fraction of the ER-Mitf protein expressed is active in the absence of ligand. Flow cytometry on the cells expressing ER and ER-Mitf (Fig. 1e) indicated that treatment of the ER-Mitf-expressing cells with 4-OHT led to a decrease in the S-phase and G2 populations and an increase in the G1 fraction compared with the control ER-expressing cells or ER-Mitf cells in the absence of 4-OHT. Thus, Mitf can control cell cycle progression.

Because the *INK4a* gene is not expressed in NIH 3T3 cells^{4,5} we examined the effect of Mitf on the expression of *p21^{Cip1}* that is required for the viability of melanocytes in culture under conditions of low serum and withdrawal of 12-O-tetradecanoylphorbol-13-acetate⁶. Addition of 4-OHT to the ER-Mitf cell line led to an increase in *p21^{Cip1}* protein (Fig. 2a) and messenger RNA (Fig. 2b) not seen in the control cells expressing the ER ligand-binding domain only. We saw no induction of *Bcl2*, a gene proposed to be a target of Mitf⁷. The mouse and human *p21^{Cip1}* promoters both contain a full consensus Mitf-binding site (T)CATGTG(A) (E2) (Fig. 2c) that includes a critical T or A residue flanking the core CATGTG E-box defined previously⁸. This arrangement is found in the M-box from the *Tyrosinase* gene, a canonical Mitf target⁹, but not adjacent to a second CATGTG motif (E1) found in the *p21^{Cip1}* promoter, nor in the proposed *Bcl2* promoter target element⁷ nor in the *P*-gene promoter that is known not to bind Mitf⁸. Binding assays *in vitro* (Fig. 2d) using a *p21^{Cip1}* E2 probe confirmed that Mitf bound the *p21^{Cip1}* E2 element as efficiently as the M-box, but bound the *Bcl2*, *p21^{Cip1}* E1 and *P*-gene E-boxes, about 15–20-fold less well. Although anti-Mitf antibody was included in the assay to yield a sharper DNA–Mitf complex, similar results were obtained in its absence (not shown). A chromatin immunoprecipitation assay (ChIP) from 501mel melanoma cells, which express both Mitf and *p21^{Cip1}*, or melan-a melanocytes, using an anti-Mitf antibody and polymerase chain reaction (PCR) primers spanning the *p21^{Cip1}*

E2 element confirmed Mitf binding to the *p21^{Cip1}* promoter *in vivo* (Fig. 2e). No signal above background was seen with non-specific IgG or with primers specific for the *HSP70* promoter as controls. Moreover, co-transfection assays revealed that Mitf could activate expression from a wild-type (WT) *p21^{Cip1}* promoter-luciferase reporter up to fivefold, but that mutation of the E2 motif severely impaired the response (Fig. 2f). Short interfering RNA (siRNA)-mediated inhibition of Mitf expression led to a significant decrease in *p21^{Cip1}* protein concentration and mRNA expression in either 501mel melanoma cells (Fig. 2g) or primary human melanocytes (Fig. 2h). Our results are consistent with *p21^{Cip1}* being a direct target for Mitf, although it is also likely that indirect regulation of *p21^{Cip1}* expression by Mitf also occurs (see below).

To determine whether the Mitf-induced G1 arrest required *p21^{Cip1}* expression we expressed ER-Mitf in RAT-1 cells in which expression of the *p21^{Cip1}* gene is silenced by promoter methylation¹⁰. Immunofluorescence (Fig. 3a) and western blotting (Fig. 3b) confirmed that ER-Mitf in the RAT-1 cells was regulated

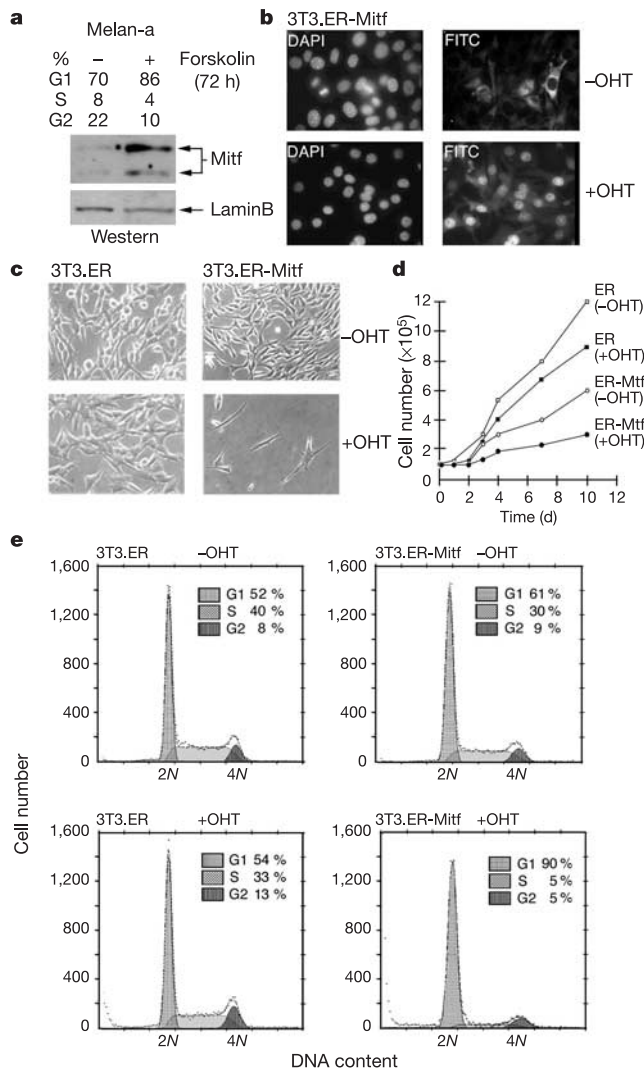


Figure 1 Mitf is a negative-regulator of cell cycle progression. **a**, Melan-a cells were treated with 20 μ M forskolin for 72 h before analysis by flow cytometry or western blotting with anti-Mitf antibody as indicated. **b**, NIH 3T3 cells stably expressing HA.ER-Mitf or HA.ER only were grown and treated with or without 4-OHT for 96 h before analysis by immunofluorescence using anti-HA antibody. **c**, Brightfield image. The cells were plated at the same density and allowed to grow for 10 days before being observed. **d**, Growth curve. **e**, Flow cytometry.

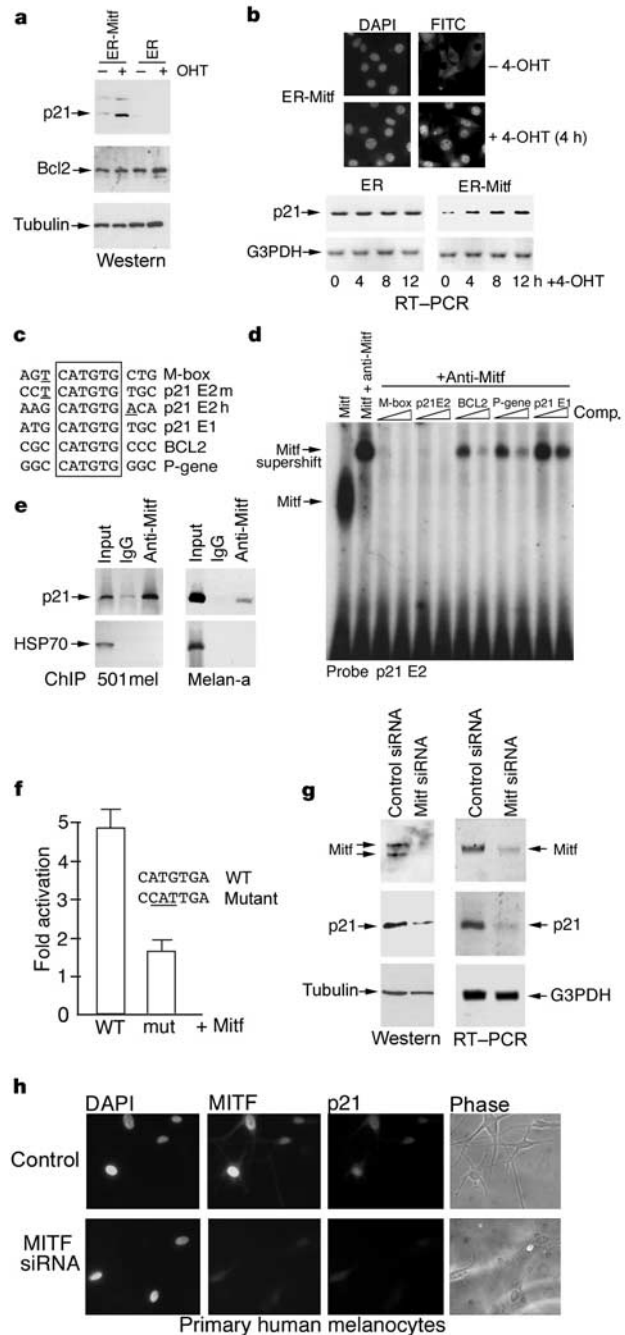


Figure 2 Mitf regulates *p21^{Cip1}* expression. **a**, Western blot, with the indicated antibodies, of NIH 3T3 cells stably expressing HA.ER-Mitf or HA.ER only grown with or without 4-OHT for 96 h. **b**, Immunofluorescence with anti-HA antibody, and RT-PCR of HA.ER-expressing NIH 3T3 cells at the indicated times after the addition of 4-OHT. **c**, M-box and E-box motifs from the indicated promoters. **d**, Band shift assay with a human *p21^{Cip1}* E2 element probe (0.2 ng), purified His-tagged Mitf and competitor oligonucleotides at 10 and 50 ng. Anti-Mitf monoclonal antibody was added where indicated. **e**, ChIP assay from 501mel or melan-a cells with the use of anti-Mitf antibody or non-specific IgG and PCR primers for the *p21^{Cip1}* or *HSP70* promoters. Quantification indicates that the signal normalized to the *HSP70* control is about fivefold enriched with the anti-Mitf antibody in comparison with IgG. **f**, Luciferase assay of cells transfected with WT or E2 site-mutated *p21^{Cip1}* promoter-luciferase reporter together with an Mitf expression vector or control vector. Results are means and s.d. **g**, Western blot with indicated antibodies or RT-PCR of 501mel melanoma cells transfected with Mitf-specific or control siRNA. **h**, Immunofluorescence, with anti-Mitf or anti-*p21^{Cip1}* antibodies, of primary human melanocytes transfected with Mitf-specific or control siRNA.

by 4-OHT and was expressed to similar levels to the NIH 3T3 cells. However, in contrast to the expression of ER-Mitf in the NIH 3T3 cells, p21^{Cip1} was not induced in the RAT-1 cells (Fig. 3c), and no substantial effect was observed on the cell cycle profile after treatment with 4-OHT (Fig. 3d), indicating that the effect of Mitf on the cell cycle requires the expression of the gene encoding p21^{Cip1}, which is implicated in G1/S checkpoint control¹¹ in fibroblasts.

The dependence of the Mitf-mediated cell cycle arrest on p21^{Cip1}

expression was confirmed by using genetically matched WT and p21^{-/-} primary mouse embryo fibroblasts (MEFs). The addition of 4-OHT to WT MEFs expressing ER-Mitf led to its accumulation in the nucleus (Fig. 3e), slow growth (Fig. 3f) and accumulation of cells in G1 (Fig. 3g), whereas no significant effect was observed in control cells expressing ER only or in p21^{-/-} MEFs expressing ER-Mitf (Fig. 3e, f, h).

Previous work has established that inactivation of Rb1 in mouse

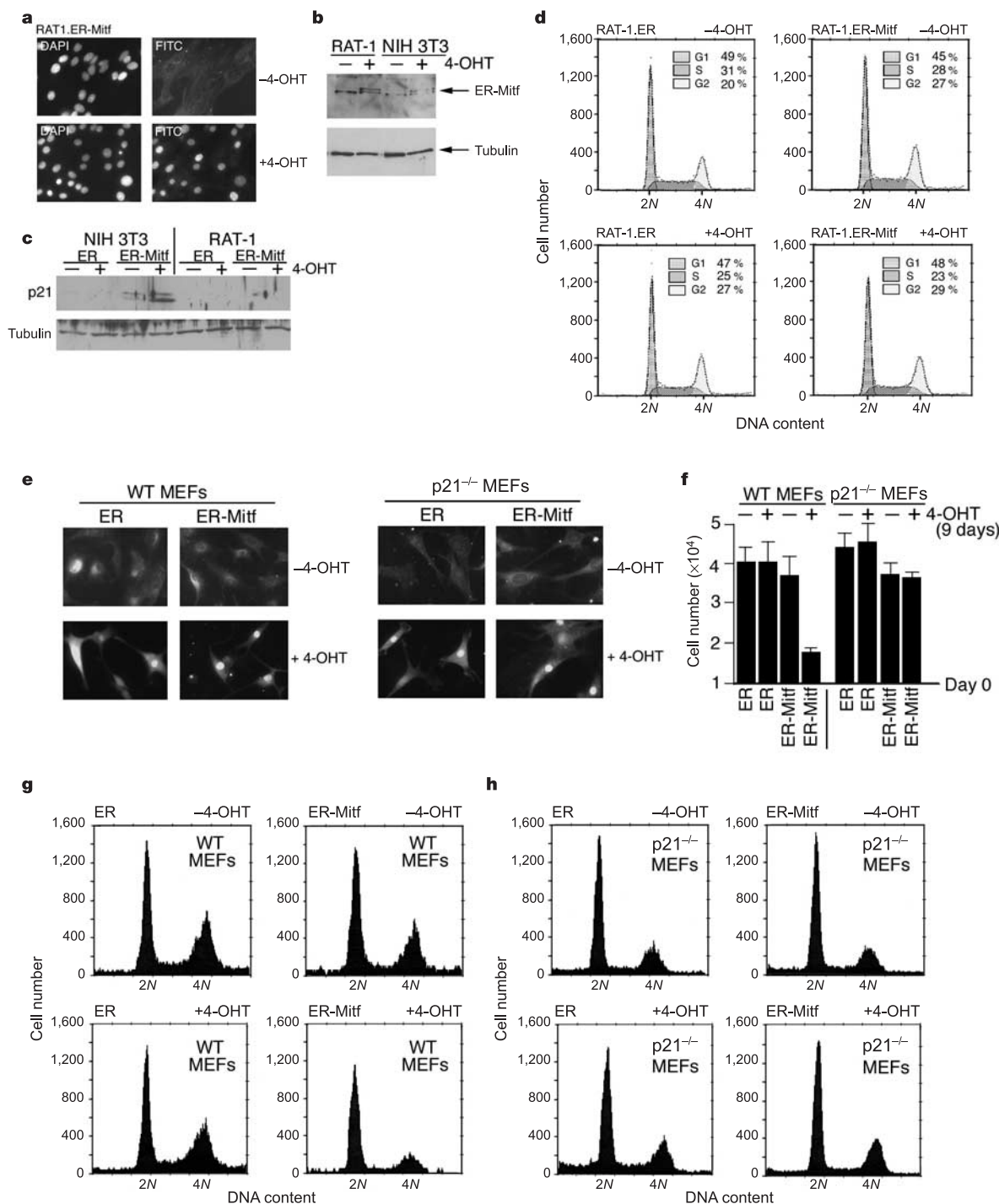


Figure 3 Mitf does not cause a cell cycle arrest in p21^{Cip1}-deficient cells. **a–d**, RAT-1 cells stably expressing HA.ER-Mitf or expressing HA.ER only were grown and treated with or without 4-OHT for 96 h and analysed by immunofluorescence (**a**), western blotting using anti-HA antibody (**b**) or anti-p21 antibody comparing ER-Mitf-expressing NIH 3T3 cells and RAT-1 cells (**c**), and flow cytometry (**d**). **e**, Immunofluorescence of WT and

p21^{-/-} MEFs expressing HA.ER or HA.ER-Mitf. **f**, Proliferation of WT and p21^{-/-} MEFs expressing HA.ER or HA.ER-Mitf. Cells (10⁴) were plated and grown in the presence or absence of 4-OHT for 9 days before being counted. Results are means and s.d. **g, h**, Flow cytometry of WT and p21^{-/-} MEFs expressing HA.ER or HA.ER-Mitf as indicated.

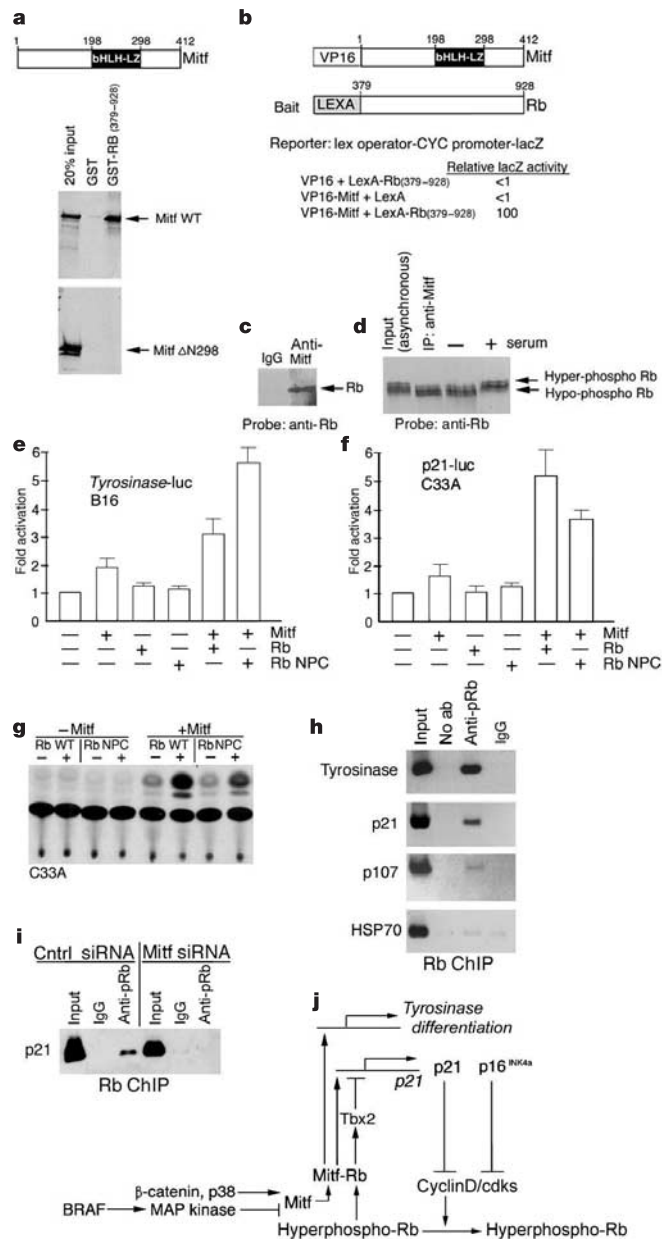


Figure 4 Mitf cooperates with Rb1. **a**, GST pull-down assay with Mitf transcribed and translated *in vitro*. **b**, Yeast two-hybrid assay with LexA-Rb1 bait with VP16-Mitf and lex-op lacZ reporter. **c**, Mitf-Rb1 co-immunoprecipitation from 501mel cells by using anti-Mitf or non-specific IgG followed by a western blot with anti-Rb1 antibody. **d**, Resolution of phosphorylated forms of Rb1. **e**, Luciferase assays using B16 cells transfected with 50 ng *Tyrosinase* promoter reporter and indicated expression vectors for Rb1, non-phosphorylatable Rb1 (NPC) or Mitf as indicated, or empty vector controls (100 ng). **f**, Luciferase assays of Rb1-negative C33A cells transfected with a p21-luciferase reporter (50 ng) and indicated expression vectors or empty vector controls (100 ng). Results in **e** and **f** are means and s.d. **g**, C33A cells were transfected with a 2 x M-box thymidine kinase-chloramphenicol acetyltransferase (CAT) reporter alone or together with indicated expression vectors. **h**, ChIP with anti-Rb1 antibody and 501mel cells together with primers spanning the *Tyrosinase*, *p21^{Cip1}* and *HSP70* promoters. **i**, Rb1 ChIP for binding to the *p21^{Cip1}* promoter by using cells treated with control (ctrl) or Mitf-specific siRNA as indicated. **j**, Model for the role of Mitf in pigment cell proliferation and differentiation. Cell cycle arrest and differentiation mediated by Mitf and Rb1 cooperation may be promoted or abrogated by signalling pathways regulating Mitf or Tbx2 expression or function. Cdk, cyclin-dependent kinases; MAP, mitogen-activated protein.

skin leads to a specific loss of melanocytes¹², implying that, like Mitf³, Rb1 is required for postnatal melanocyte survival. We therefore examined whether Mitf might interact with Rb1. Glutathione S-transferase (GST) pull-down assays (Fig. 4a) using the Rb1 carboxy-terminal region and pocket domain (amino acids 379–928) indicated that Rb1 could interact with full-length Mitf transcribed and translated *in vitro*, which is consistent with our previous observation¹⁴, but not a deletion mutant lacking the amino terminus and DNA-binding domain. No interaction was observed with GST only. A similar result was obtained in a two-hybrid assay with yeast expressing VP16-Mitf together with a bait comprising Rb1 residues 379–928 fused to the LexA DNA-binding protein and a Lex operator–CYC promoter *LacZ* reporter (Fig. 4b), although the interaction was decreased tenfold compared with that obtained using E1A (not shown). The potential interaction was confirmed by immunoprecipitation of endogenous Mitf with an anti-Mitf antibody and western blotting with an anti-Rb1 antibody (Fig. 4c). Separation of the phosphorylated forms of Rb1 and analysis by western blot suggested that Mitf can specifically interact with hypophosphorylated Rb1 using extracts from serum-starved or stimulated cells to provide a source of hypophosphorylated or hyperphosphorylated Rb1 as markers (Fig. 4d), although it is difficult to establish whether Mitf forms a stable complex with Rb1 or whether the interactions detected are transient.

To determine whether Mitf and Rb1 could act cooperatively we transfected B16 melanoma cells with a *Tyrosinase*–luciferase reporter together with limiting amounts of vectors expressing Mitf, WT Rb1 or a constitutively active version of Rb1 in which 14 phosphorylation sites have been mutated in the N terminus, pocket and C terminus (Rb1 NPC)¹⁵. The results (Fig. 4e) indicate that whereas Mitf could activate the reporter weakly in the absence of Rb1, the co-expression of either WT or Rb1 NPC led to an enhanced activation. Note that because B16 cells contain both endogenous Mitf and Rb1, transfection of high levels of either Mitf or Rb1 expression vectors alone can activate the reporter (not shown) but the low levels of DNA transfected into the cells in this assay allow the cooperative effects of Mitf and Rb1 to be seen. Similar cooperative effects of Rb1 and Mitf were also observed on a *p21^{Cip1}*–luciferase reporter in Rb1-negative C33a cells (Fig. 4f) or with a reporter in which chloramphenicol acetyltransferase expression is driven by two Mitf-binding sites (M-box) placed upstream of a thymidine kinase promoter (Fig. 4g). With the p107 promoter as a positive control¹⁶, ChIPs from 501mel melanoma cells with the use of anti-Rb1 antibody revealed that Rb1 was present at both the Mitf-dependent *Tyrosinase* and *p21^{Cip1}* promoters but not at the Mitf-independent *HSP70* promoter (Fig. 4h), and that siRNA-mediated depletion of Mitf led to a decreased Rb1 ChIP at the *p21^{Cip1}* promoter (Fig. 4i).

The fact that Mitf can induce a cell cycle arrest in cells lacking a functional *INK4a* gene, but not in *p21^{Cip1}*-deficient cells, indicates that Mitf regulates cell cycle progression through *p21^{Cip1}*. As shown in Fig. 4j, the ability of Mitf to cooperate with Rb1 and activate *p21^{Cip1}* expression could contribute to a positive feedback loop leading to cell cycle exit and differentiation reminiscent of the role of CCAAT-enhancer-binding protein, Rb1 and *p21^{Cip1}* in adipocyte differentiation¹⁷. Moreover, even in cycling cells, for example in hair follicle melanocytes, Mitf-mediated regulation of cell cycle progression could also contribute to cell survival, allowing cells to progress through the cell cycle only when appropriate survival signals, such as that provided by Kit receptor tyrosinase kinase¹⁸, are present. Thus Mitf might act to integrate a variety of extracellular signals with cell division, a notion consistent with the fact that cultured *p21^{Cip1}*-deficient melanocytes exhibit a survival defect⁶. The physiological relevance of our cell-based studies is underpinned by genetic evidence that Mitf regulates pigment cell proliferation. Thus, in addition to indirect evidence discussed previously¹, Mitf heterozygous mice (*Mitf/Mitf^{mi}*) are characterized by increased

proliferation of melanoblasts¹⁹; perhaps more strikingly, the hypo-proliferation defect observed in the developing neuroretina of *Chx10* mutant mice is largely corrected if *Mitf* is also lost²⁰. Together with our results, the evidence is strongly indicative that *Mitf* can negatively regulate cell cycle progression *in vivo*.

About 70% of melanomas are characterized by constitutive activation of BRAF²¹, which would act to increase *Mitf* turnover through the phosphorylation of mitogen-activated protein kinase and potentially decrease the impact of *Mitf* on cell cycle control¹. Indeed, *p21^{Cip1}* expression levels have been shown to be inversely correlated with melanoma progression²². However, it is important to stress that although cooperation between *Mitf* and Rb1 and the regulation of *p21^{Cip1}* expression might regulate cell cycle progression during melanocyte differentiation or in melanoma, its contribution to the survival of melanoblasts during development is less obvious, because *p21^{Cip1}*-deficient mice are pigmented¹¹ whereas *Mitf*^{-/-} animals lack all melanocytes²³. Indeed, the ability of *Mitf* to regulate *p21^{Cip1}* and the cell cycle is likely to be highly complex, with many factors contributing directly or indirectly to the regulation of *p21^{Cip1}* expression by *Mitf*. For example, *Mitf* can also activate *Tbx2* expression²⁴ but *Tbx2* can both repress the *p21^{Cip1}* promoter and suppress senescence^{25,26}, whereas preliminary results (not shown) also indicate that in some cells *Mitf* might be essential for proliferation, perhaps indicating that too much or too little *Mitf* is detrimental to cell cycle progression. Nevertheless, the results presented here identify *Mitf* as a novel activator of *p21^{Cip1}* expression and an important regulator of cell cycle progression. □

Methods

Generation and growth of cell lines

All cell lines were grown as described²⁴ but with primary human melanocytes being cultured in melanocyte growth medium (Gentaur). B16 are *INK4a*-null Melan-a cells, 501mel cells do not express p16^{INK4a}, and C33A, RAT-1 and MEFs are *INK4a* WT. Throughout this paper *Mitf* refers to the *Mitf*-M(+) isoform and the cDNA used was of mouse origin. Virus-producing cell lines were produced by transfection of Psi2 cells with pBabe. Haemagglutinin epitope (HA)ER²⁷ or pBabe.HA.ER-*Mitf* expression vectors, expressing cells selected with puromycin, and viral supernatants used to infect target cells as described²⁵. For the RAT-1 and NIH 3T3 cells individual clones were isolated and expanded under selection conditions. ER fusion proteins were activated by the addition of 4-OHT to a final concentration of 300 nM. For the MEFs, puromycin-resistant cells were pooled before analysis.

siRNA-mediated downregulation of *Mitf* was achieved with the following *Mitf*-specific sequences: 5'-r(AGCAGUACCUUUCUACAC)d(TT)-3' and 5'-r(GUGGUAGAAAGG UACUGCU)d(TT)-3'. The control siRNA sequences used were 5'-r(UUCUCCGAACGU GUCACGU)d(TT)-3' and 5'-r(ACGUGACACGUUCGAGAA)d(TT)-3'.

PCR with reverse transcription (RT-PCR) was performed as described²⁸ with primers specific for *Mitf* (5'-ATGCTGCAAATGCTAGATAA-3' and 5'-CAATCAGGTTGTGATT GTCC-3'), *p21^{Cip1}* (5'-CTAGGGCTGCCAA-3' and 5'-TCAGGTTTCTCTTTCAGC-3') and *G3PDH* (5'-CCAACTGCTTAGCCCCCTGGCCAAG-3' and 5'-CTCCTTGA GGCCATGTAGGCCATG-3').

After siRNA treatment for 3 days, cells were assayed by western blotting, immunofluorescence or RT-PCR. Luciferase assays were performed as described²⁴ and assayed 2 days after transfection.

Flow cytometry

Cells were harvested and fixed as a single cell suspension in methanol at -20 °C. Cells were recovered by centrifugation, then stained with propidium iodide and subjected to analysis with a Beckman Coulter EpicsXL flow cytometer.

Western blot analysis

Whole cell extracts were prepared as described²⁵. The primary antibodies used were rabbit polyclonal anti-p21 (C-19) and anti-Bcl2 (Santa Cruz Biotechnology), anti-HA mouse monoclonal and anti- α -tubulin rabbit polyclonal (Sigma), anti-*Mitf* mouse monoclonal (C5; Neomarkers) and mouse anti-Rb1 monoclonal (Pharmingen).

Yeast strains, media and methods

Standard methods for yeast culture and transformation were followed and two-hybrid assays were performed as described previously²⁹. The WT VP16-*Mitf* expression vector has been described previously⁸. The Rb1 sequences encoding amino acids 379–928 were cloned as a *Bam*HI fragment into the *Bam*HI site of the LexA expression vector²⁹. All assays were performed in strain DY1641 containing an integrated lex op-CYC-lacZ reporter (MATa, *ade2-1*, *ura3-52*, *his3-11.15*, *trp1-1*, *can1-100*, *leu2-3,112*, *URA3::lexopx8-lacZ*).

Immunofluorescence microscopy

Immunofluorescence was performed as described²⁵ with a 1:100 dilution of monoclonal anti-HA (Clone HA-7; Sigma), monoclonal anti-*Mitf* (Neomarkers) or anti-p21 (Santa Cruz) antibodies and with a 1:100 dilution of appropriate secondary antibodies (Vector Laboratories).

GST pull-down assays and immunoprecipitations

GST fusion proteins were expressed in *Escherichia coli* strain BL21(DE3) pLysS and purified as described previously⁸. GST pull-down assays were performed as described¹⁴. For immunoprecipitations, extracts were prepared from 10-cm dishes of 501mel cells by lysis in 1.5 ml RIPA buffer (50 mM Tris-HCl pH 8.0, 0.1% SDS, 0.01% sodium deoxycholate, 150 mM NaCl, 1% Triton X-100, protease inhibitor cocktail (Roche Molecular Biochemicals)) for 15 min on ice and clarified by centrifugation. Extracts were precleared with 50 μ l Protein A-Sepharose for 2 h at 4 °C and the beads were then removed by centrifugation. Anti-*Mitf* monoclonal antibody (10 μ l) was added and incubated for 16 h at 4 °C with rotation. Protein A-Sepharose (50 μ l) was then added and incubated for 3 h at 4 °C with rotation to capture the immune complexes. Beads were pelleted and washed five times with RIPA buffer, then bound proteins were eluted by heating in 20 μ l SDS-polyacrylamide-gel electrophoresis loading buffer, analysed by SDS-polyacrylamide-gel electrophoresis and western blotting with anti-Rb1 monoclonal antibody. To separate the different phosphorylated forms of Rb1 we used a polyacrylamide/bisacrylamide ratio of 200:1.

Chromatin immunoprecipitation

ChIP assays were performed as described previously for *Mitf*³⁰ with 10 μ g anti-*Mitf* mouse monoclonal antibody or 10 μ g non-specific IgG (Bio-Rad). For the Rb1 ChIPs we used the mouse anti-human Rb1 monoclonal (Pharmingen) and the protocol described by P. Farnham (<http://genomecenter.ucdavis.edu/farnham/farnham/protocols/chips.html>). The DNA recovered was subjected to amplification by PCR before analysis by agarose-gel electrophoresis. The primers used for the PCRs were the human *p21^{Cip1}* promoter region (5'-GGGTGTAGGGAGATTGGTCAATG-3' and 5'-CCGCTGACCCACTCTGGCA GGC-3'), the *HSP70* promoter (5'-CTCCAGTGAATCCAGAAAGACTCT-3' and 5'-TGGGACAAACGGGAGTCACTCTC-3') and the *Tyrosinase* promoter (5'-GCTCTATT CCTGACACTACCTCTC-3' and 5'-CAAGTGCTGCAGAACTGGCTAATTG-3').

DNA-binding assays

Electrophoretic mobility-shift assays were performed with purified His-tagged *Mitf* and ³²P-labelled oligonucleotide *p21^{Cip1}* E2 probe as described previously²⁴. The sequences for probe and competitor DNA were as follows: *p21^{Cip1}* E2, 5'-CTAGATTCGCGAAGCAT GTGACAAATCAACA-3'; M-box, 5'-CTAGACAGTGGGGAGGGAGTCAATGTGCTGCC TAGTT-3'; *BCL2*, 5'-CTAGACCCCGCGGCCATGTGCCCCCGCGGT-3'; *p21^{Cip1}* E1, 5'-CTAGAGATCTGATGATGTGTGCTTG-3'; and *P-gene*, 5'-CTAGAAGCTGT GGGCGGGGCCATGTGGCAGCTTTGTT-3'.

Received 24 May; accepted 9 December 2004; doi:10.1038/nature03269.

- Vance, K. W. & Goding, C. R. The transcription network regulating melanocyte development and melanoma. *Pigment Cell Res.* **17**, 318–325 (2004).
- Bertolotto, C. *et al.* Microphthalmia gene product as a signal transducer in cAMP-induced differentiation of melanocytes. *J. Cell Biol.* **142**, 827–835 (1998).
- Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G. & Evan, G. I. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res.* **23**, 1686–1690 (1995).
- Quelle, D. E. *et al.* Cloning and characterization of murine p16INK4a and p15INK4b genes. *Oncogene* **11**, 635–645 (1995).
- Quelle, D. E., Zindy, F., Ashmun, R. A. & Sherr, C. J. Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell* **83**, 993–1000 (1995).
- Halaban, R., Cheng, E., Zhang, Y., Mandigo, C. E. & Miglarese, M. R. Release of cell cycle constraints in mouse melanocytes by overexpressed mutant E2F1_{E132}, but not by deletion of p16^{INK4a} or p21^{WAF1/CIP1}. *Oncogene* **16**, 2489–2501 (1998).
- McGill, G. G. *et al.* Bcl2 regulation by the melanocyte master regulator *Mitf* modulates lineage survival and melanoma cell viability. *Cell* **109**, 707–718 (2002).
- Aksan, I. & Goding, C. R. Targeting the microphthalmia basic helix-loop-helix-leucine zipper transcription factor to a subset of E-box elements *in vitro* and *in vivo*. *Mol. Cell. Biol.* **18**, 6930–6938 (1998).
- Bentley, N. J., Eisen, T. & Goding, C. R. Melanocyte-specific expression of the human tyrosinase promoter: activation by the microphthalmia gene product and role of the initiator. *Mol. Cell. Biol.* **14**, 7996–8006 (1994).
- Allan, L. A., Duhig, T., Read, M. & Fried, M. The p21(WAF1/CIP1) promoter is methylated in Rat-1 cells: stable restoration of p53-dependent p21(WAF1/CIP1) expression after transfection of a genomic clone containing the p21(WAF1/CIP1) gene. *Mol. Cell. Biol.* **20**, 1291–1298 (2000).
- Deng, C., Zhang, P., Harper, J. W., Elledge, S. J. & Leder, P. Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control. *Cell* **82**, 675–684 (1995).
- Yu, B. D. *et al.* Distinct and nonoverlapping roles for pRB and cyclin D: cyclin-dependent kinases 4/6 activity in melanocyte survival. *Proc. Natl Acad. Sci. USA* **100**, 14881–14886 (2003).
- Lerner, A. B. *et al.* A mouse model for vitiligo. *J. Invest. Dermatol.* **87**, 299–304 (1986).
- Yavuzer, U. *et al.* The microphthalmia gene product interacts with the retinoblastoma protein *in vitro* and is a target for deregulation of melanocyte-specific transcription. *Oncogene* **10**, 123–134 (1995).
- Chew, Y. P., Ellis, M., Wilkie, S. & Mittnacht, S. pRB phosphorylation mutants reveal role of pRB in regulating S phase completion by a mechanism independent of E2F. *Oncogene* **17**, 2177–2186 (1998).
- Takahashi, Y., Rayman, J. B. & Dynlacht, B. D. Analysis of promoter binding by the E2F and pRB families *in vivo*: distinct E2F proteins mediate activation and repression. *Genes Dev.* **14**, 804–816 (2000).

17. Umek, R. M., Friedman, A. D. & McKnight, S. L. CCAAT-enhancer binding protein: a component of a differentiation switch. *Science* **251**, 288–292 (1991).
18. Okura, M., Maeda, H., Nishikawa, S. & Mizoguchi, M. Effects of monoclonal anti-c-kit antibody (ACK2) on melanocytes in newborn mice. *J. Invest. Dermatol.* **105**, 322–328 (1995).
19. Hornyak, T. J., Hayes, D. J., Chiu, L. Y. & Ziff, E. B. Transcription factors in melanocyte development: distinct roles for Pax-3 and Mitf. *Mech. Dev.* **101**, 47–59 (2001).
20. Konyukhov, B. V. & Sazhina, M. V. Interaction of the genes of ocular retardation and microphthalmia in mice. *Folia (Praha)* **12**, 116–123 (1966).
21. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).
22. Jiang, H. *et al.* The melanoma differentiation-associated gene mda-6, which encodes the cyclin-dependent kinase inhibitor p21, is differentially expressed during growth, differentiation and progression in human melanoma cells. *Oncogene* **10**, 1855–1864 (1995).
23. Hodgkinson, C. A. *et al.* Mutations at the mouse *microphthalmia* locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell* **74**, 395–404 (1993).
24. Carreira, S., Liu, B. & Goding, C. R. The gene encoding the T-box transcription factor Tbx2 is a target for the microphthalmia-associated transcription factor in melanocytes. *J. Biol. Chem.* **275**, 21920–21927 (2000).
25. Prince, S., Carreira, S., Vance, K. W., Abrahams, A. & Goding, C. R. Tbx2 directly represses the expression of the p21(WAF1) cyclin-dependent kinase inhibitor. *Cancer Res.* **64**, 1669–1674 (2004).
26. Jacobs, J. J. *et al.* Senescence bypass screen identifies TBX2, which represses cdkn2a (p19ARF) and is amplified in a subset of human breast cancers. *Nature Genet.* **26**, 291–299 (2000).
27. Vigo, E. *et al.* CDC25A phosphatase is a target of E2F and is required for efficient E2F-induced S phase. *Mol. Cell. Biol.* **19**, 6379–6395 (1999).
28. Carreira, S., Dexter, T. J., Yavuzer, U., Easty, D. J. & Goding, C. R. Brachyury-related transcription factor Tbx2 and repression of the melanocyte-specific TRP-1 promoter. *Mol. Cell. Biol.* **18**, 5099–5108 (1998).
29. Hirst, K., Fisher, F., McAndrew, P. C. & Goding, C. R. The transcription factor, the Cdk, its cyclin and their regulator: directing the transcriptional response to a nutritional signal. *EMBO J.* **13**, 5410–5420 (1994).
30. Galibert, M. D., Carreira, S. & Goding, C. R. The Usf-1 transcription factor is a novel target for the stress-responsive p38 kinase and mediates UV-induced tyrosinase expression. *EMBO J.* **20**, 5022–5031 (2001).

Acknowledgements We thank M. Serrano for the WT and p21-null MEFs; S. Mittnacht for the Rb1 expression vectors and the C33a cells; B. Vogelstein for the *p21^{Cip1}* promoter; K. Helin for the HA,ER-expression vector; D. Stillman for yeast reporter strain DY1641; C. Wellbrock, R. Marais, R. Ballotti and C. Bertolotto for communication of unpublished results; and H. Arnheiter for discussions. This work was supported by Marie Curie Cancer Care, the Association for International Cancer Research and a European Union Marie Curie fellowship to M.-D.G.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.R.G. (c.goding@mcrci.ac.uk).

Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs

Lee P. Lim¹, Nelson C. Lau², Philip Garrett-Engele¹, Andrew Grimson², Janell M. Schelter¹, John Castle¹, David P. Bartel², Peter S. Linsley¹ & Jason M. Johnson¹

¹Rosetta Inpharmatics (wholly owned subsidiary of Merck and Co.), 401 Terry Avenue N, Seattle, Washington 98109, USA

²Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

MicroRNAs (miRNAs) are a class of noncoding RNAs that post-transcriptionally regulate gene expression in plants and animals^{1,2}. To investigate the influence of miRNAs on transcript levels, we transfected miRNAs into human cells and used microarrays to examine changes in the messenger RNA profile. Here we show that delivering miR-124 causes the expression profile to shift towards that of brain, the organ in which miR-124 is preferentially expressed, whereas delivering miR-1 shifts the profile towards that of muscle, where miR-1 is preferentially expressed. In each case, about 100 messages were downregulated after 12 h. The 3' untranslated regions of these messages had a

significant propensity to pair to the 5' region of the miRNA, as expected if many of these messages are the direct targets of the miRNAs³. Our results suggest that metazoan miRNAs can reduce the levels of many of their target transcripts, not just the amount of protein deriving from these transcripts. Moreover, miR-1 and miR-124, and presumably other tissue-specific miRNAs, seem to downregulate a far greater number of targets than previously appreciated, thereby helping to define tissue-specific gene expression in humans.

Gene expression analysis has helped find targets of an over-expressed plant miRNA⁴, but such an approach has not been reported in animals, where, in contrast to plants, miRNAs are believed to act mainly through translational repression rather than messenger RNA (mRNA) cleavage^{1,2}. Nonetheless, a microarray analysis showed that unintentional reduction of transcript levels can be observed following small interfering RNA (siRNA) transfection of HeLa cells, and it was observed that several of these downregulated, off-target transcripts contained sites of partial complementarity to the siRNA⁵, reminiscent of those seen between metazoan miRNAs and their targets. This observation, together with the documented overlap between the potential activities of siRNAs and metazoan miRNAs^{6–9} led us to hypothesize that the off-target effects of siRNAs are due to miRNA-like activity, and therefore using microarrays to screen for changes in gene expression following transfection of human miRNAs would shed light on natural miRNA functions.

We focused on two miRNAs noted for their tissue-specificity in mammals: miR-1 and miR-124. miR-1 is preferentially expressed in heart and skeletal muscle and miR-124 is preferentially expressed in brain^{10,11} (Supplementary Fig. 1). miR-1 or miR-124 RNA duplexes were transfected into HeLa cells, and mRNA was purified and profiled on microarrays (Fig. 1a). Filtering the expression profiles for genes characterized by the LocusLink database¹² that were significantly downregulated ($P < 0.001$) at both 12 and 24 h, gave sets of 96 and 174 annotated genes downregulated by miR-1 and miR-124, respectively (Supplementary Tables 1 and 2).

One striking feature of the 174 genes downregulated by miR-124 is that these genes are generally expressed at lower levels in the brain than in other tissues of the body. In a sense, transfecting this brain miRNA into HeLa cells shifted HeLa gene expression towards that of brain. This phenomenon is illustrated by graphing relative gene expression in cerebral cortex (Fig. 1b, c), using as reference the expression levels of 10,000 LocusLink genes across 46 human tissues¹³. An index of relative gene expression for each tissue was created by first ranking the expression levels of each gene over all tissues, whereupon the gene ranks for the tissue of interest were plotted in a histogram. For example, the left-most column (rank 1) in the distribution of gene ranks for cerebral cortex (Fig. 1b) represents the number of genes that are expressed more highly in cerebral cortex than in any other tissue, whereas the right-most column (rank 46) represents the genes that are expressed less in cerebral cortex than in any other tissue. If we were to randomly select a set of genes from the microarray and examine their cerebral cortex rankings, the distribution of rankings would approximate this background distribution. The distribution of cerebral cortex rankings for the miR-124 downregulated genes (Fig. 1c) clearly differs from the background distribution: the 174 genes downregulated by miR-124 transfection are significantly enriched for genes that are expressed at lower levels in cerebral cortex relative to other tissues.

When this analysis was extended to all 46 tissues, significant differences from background were observed only for brain tissues (Fig. 1d), mirroring the known tissue distribution of miR-124. When the miR-1 results were analysed in the same manner, the most significant P values were not observed in brain but instead were observed for heart and skeletal muscle (Fig. 1e), the two tissues known to express miR-1. Therefore, the *in vivo* expression specifi-

cities of miR-1 and miR-124 are coded in the identities of the genes that these miRNAs downregulate *in vitro*, indicating that these experiments are capturing aspects of the *in vivo* activity of miRNAs.

The shift towards brain-like or muscle-like expression need not result from the miRNA interacting directly with the downregulated messages; in principle, it could be a secondary effect of repressing key regulatory genes. The most compelling evidence that the miRNA-downregulated messages are direct targets comes from analysis of sequence motifs over-represented in these messages. The motif discovery tool MEME¹⁴ was used to computationally search the 3' untranslated regions (UTRs) of the downregulated transcripts. For the miR-1 set, highly significant motifs were found, with a length of six or more bases (Fig. 2a). The six-nucleotide (6-nt) consensus sequence, CAUUCU, is exactly complementary to positions 2–7 from the 5' end of miR-1, and was found in 88% of the annotated UTRs. For longer motif sizes, sequences were found that extend the complementarity with miR-1. A sequence logo representation of the motif of length 10 illustrates the important

contributions of positions 1–8, and especially 2–7. Similar results were seen for miR-124 (Fig. 2b). Another analytical method—comparison of hexamer frequencies in downregulated UTRs versus all UTRs—verified the MEME findings (Supplementary Table 3). These results imply that knockdown of the transcripts is caused mainly by direct binding of the transfected miRNAs to the 3' UTRs, with binding at the 5' end of the miRNA being particularly vital.

The 5' region, and particularly seed positions 2–8, is the most conserved region of metazoan miRNAs^{3,15} and has long been thought to play a key role in target recognition^{16,17}, a hypothesis recently supported both computationally and experimentally^{3,18}. Here, the demonstration that the miRNA seed region can be identified without prior knowledge of the miRNA sequence provides further experimental support for this concept, and is particularly interesting in that our experiments detected reduced mRNA levels: previously, pairing to the 5' region without additional pairing to the remainder of the miRNA had been associated with inhibition of productive translation but not reduced message levels¹⁸.

To test the importance of the 5' region for mRNA knockdown, two mutant sequences of miR-124 were designed by switching bases at positions 5 and 6 (124mut5-6) or positions 9 and 10 (124mut9-10). When RNA from HeLa cells transfected with these mutant sequences was analysed as above, the 124mut9-10 duplex gave a downregulated signature 89% shared with that of the wild type, whereas the downregulated signature for 124mut5-6 was almost completely distinct (Fig. 3a). When compared against the human tissue indices, the 124mut9-10 signature highlighted brain tissues, whereas the 124mut5-6 signature did not (Supplementary Fig. 2). These results demonstrate that, for the mRNA knockdown observed in this assay, positions 5 and 6 are much more crucial than positions 9 and 10. Additional experiments using chimaeric miRNAs showed that the 5' ends of the miRNAs were sufficient for generating tissue-specific signatures (Fig. 3b; Supplementary Fig. 2).

To further establish the link between repression and the presence of seed matches within the 3' UTR, wild-type and mutant 3' UTR segments from eight genes downregulated by miR-1 (Fig. 4a) and two genes downregulated by miR-124 (Fig. 4b) were placed into an established luciferase reporter system⁷. When cotransfected with the cognate miRNA, six of the ten wild-type reporters exhibited significant ($P < 0.001$) repression relative to the corresponding

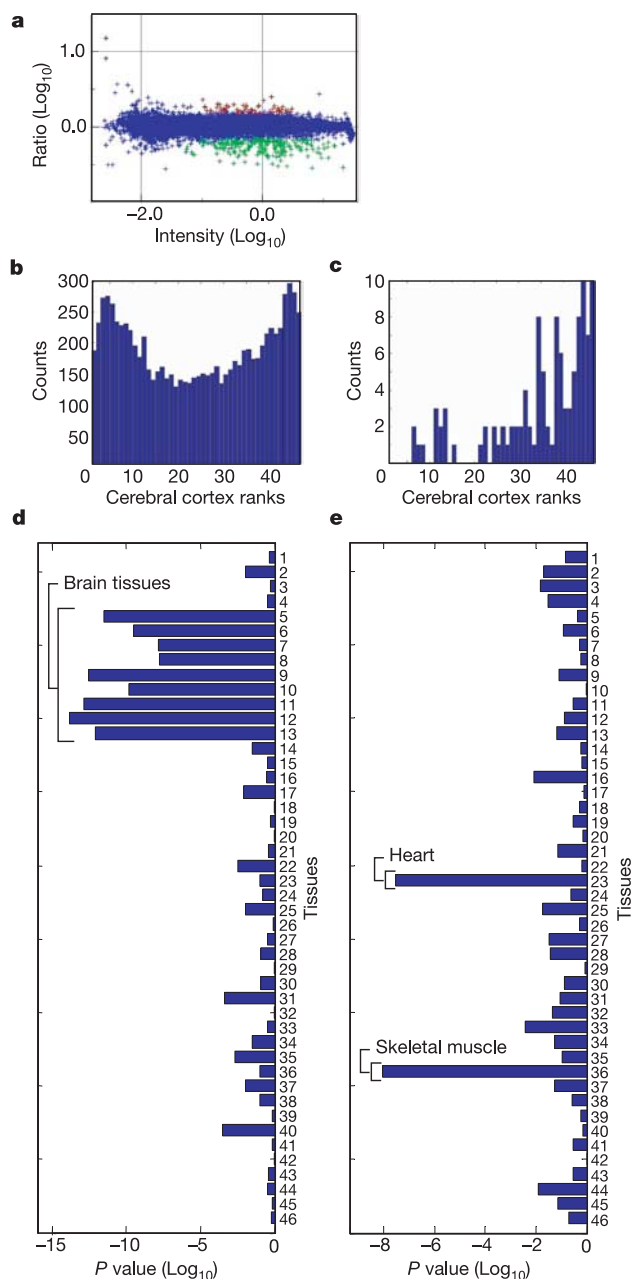


Figure 1 Tissue-specific gene expression rankings for downregulated genes.

a, Microarray signature 12 h after miR-124 transfection, plotted as expression ratio (relative to mock transfection) versus fluorescence intensity. Significantly downregulated probes ($P < 0.001$) are in green; upregulated probes in red. **b**, Background cerebral cortex rankings. The expression levels of each gene were sorted over 46 human tissues. The rankings for cerebral cortex are shown for LocusLink genes represented both on the chip and in the tissue survey. **c**, Cerebral cortex rankings for the set of genes downregulated at both 12 and 24 h following transfection of miR-124. Ranks were compiled as in **b**, and when compared with the background set were found to significantly differ using a one-sided Kolmogorov–Smirnov test ($P = 3 \times 10^{-13}$). **d**, **e**, Log (base 10) P values are plotted for each of the 46 tissues for the miR-124 (**d**) and miR-1 (**e**) downregulated sets. Two-tailed and Mann–Whitney tests gave similar results. Tissue key: 1, adrenal cortex; 2, adrenal medulla; 3, bladder; 4, bone marrow; 5, brain; 6, brain: amygdala; 7, brain: caudate nucleus; 8, brain: cerebellum; 9, brain: cerebral cortex; 10, brain: fetal; 11, brain: hippocampus; 12, brain: postcentral gyrus; 13, brain: thalamus; 14, Burkitt's lymphoma (Daudi); 15, promyelocytic leukemia (HL-60); 16, chronic myelogenous leukemia (K562); 17, lymphoblastic leukemia (MOLT-4); 18, colorectal adenocarcinoma (SW480); 19, colon: descending; 20, colon: transverse; 21, duodenum; 22, epididymus; 23, heart; 24, ileum; 25, jejunum; 26, kidney: fetal; 27, liver; 28, liver: fetal; 29, lung; 30, lung: fetal; 31, lymph node; 32, placenta; 33, prostate; 34, retina; 35, salivary gland; 36, skeletal muscle; 37, spinal cord; 38, spleen; 39, stomach; 40, testis; 41, thymus; 42, thyroid; 43, tonsil; 44, trachea; 45, uterus; 46, uterus: corpus.

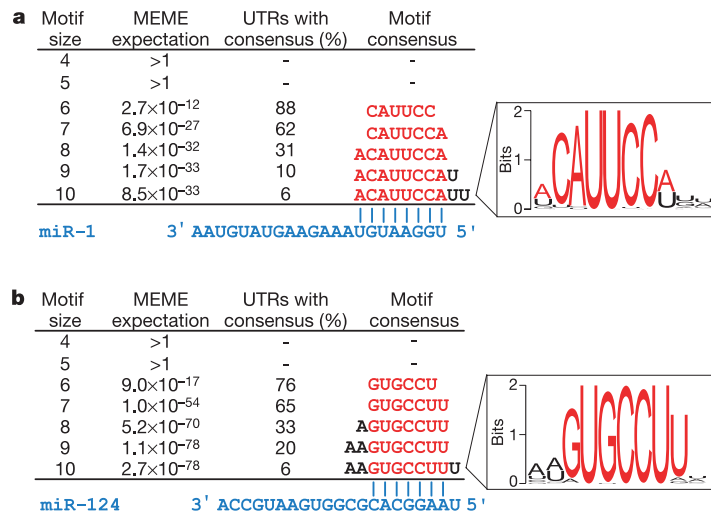


Figure 2 Over-represented motifs in the 3' UTRs of downregulated genes. **a**, **b**, miR-1 (**a**) and miR-124 (**b**) downregulated genes. Motif nucleotides that were complementary to the transfected miRNA are shown in red; base-pairing to the miRNA sequence is shown in blue. The percentage of downregulated UTRs that contained an exact match to the consensus is also shown. A sequence logo of the weight matrix that MEME constructed for the motif of size 10 is shown at the right: for a given position in the matrix, the combined

height of the bases represents the information content at that position, whereas the relative heights of the individual bases represent the frequency of that base at that position. The MEME expectation is an estimate of the number of expected motifs that would have occurred by chance at the strength and frequency of the observed motif. The lack of significant MEME motifs of length five and smaller reflects the ease at which motifs of this length can be found by chance.

constructs with mutant seed matches. Cotransfection of the non-cognate miRNA typically had no effect. Although in measuring luciferase activity these data do not distinguish between decreased mRNA levels and decreased productive translation, these data support the conclusion that pairing to the miRNA seed region contributes directly to mRNA repression. Nevertheless, 3' UTR matches to the consensus motifs of Fig. 3 are still not sufficient to accurately predict downregulation on a genome-wide level. Analysis of data from further microarray experiments may reveal additional determinants specifying which transcripts are subject to downregulation. Until improved computational methods are developed, microarray experiments should be a valuable tool for helping to identify which of the many messages with seed matches are most susceptible to miRNA-mediated knockdown.

We also searched the upregulated genes and the coding regions of downregulated genes for enrichment of hexamers complementary to the transfected miRNA. No significant enrichment in the upregulated transcripts was detected, supporting the conclusion that metazoan miRNAs are predominantly negative regulators of gene expression³. In the coding regions of the miR-1 and miR-124 downregulated messages, we did detect enrichment for the words CAUUCC and UGCCUU, respectively. However, enrichment of these words was far more pronounced in the 3' UTR (Supplementary Fig. 3), supporting the idea that miRNAs interact more frequently or more effectively with 3' UTRs than with other regions of mRNAs.

The strongest evidence that the transfection experiments described here are identifying miRNA targets comes from the observation that miRNA tissue specificity can be decoded from the expression profiles of human tissues: specifically, the set of HeLa genes downregulated by a miRNA are enriched for genes that are lowly expressed within the tissue where the miRNA is known to be naturally expressed. The simplest interpretation is that these genes downregulated in HeLa cells are predominantly biological targets of the miRNAs in human muscle (miR-1) or brain (miR-124). In support of the biological importance of these interactions, many of the potential miRNA binding sites are in evolutionarily conserved UTR regions. For example, for the transcripts in which 3' UTRs were annotated both in human and mouse, 44% of the UGCCUU sites in the miR-124-downregulated human UTRs could be aligned

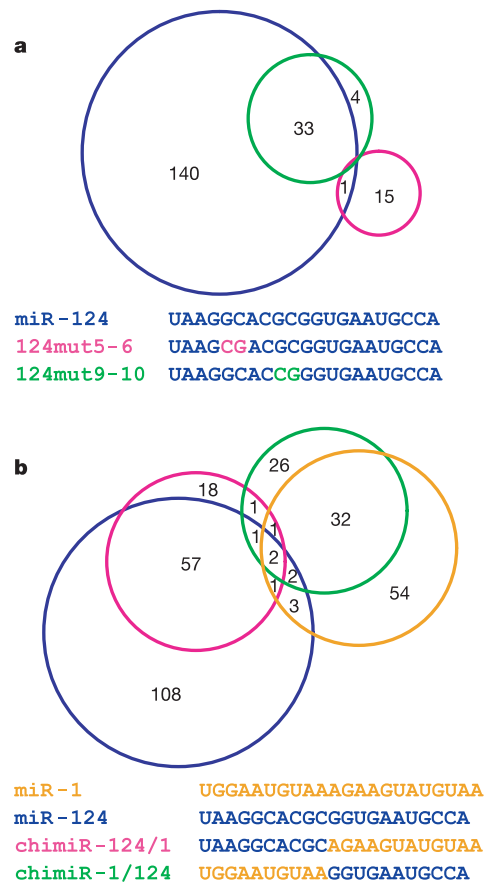


Figure 3 Microarray analysis of the effects of miRNA mutations. **a**, Sequences of wild-type and mutant miR-124 that were transfected into HeLa cells, with a Venn diagram of overlap of wild-type (blue), 124mut5-6 (pink) and 124mut9-10 (green) downregulated signatures. The smaller signatures of 124mut5-6 and 124mut9-10, when compared with wild type, were probably due to variability in transfection efficiencies. **b**, Sequences of wild-type and chimaeric miRNAs transfected into HeLa cells, with a Venn diagram of overlap of miR-1 (orange), miR-124 (blue), chimiR-124/1 (pink) and chimiR-1/124 (green) downregulated signatures. Owing to geometric constraints, a one-gene overlap between miR-1 and miR-124/1 is not pictured.

with a UGCCUU site in mouse, as opposed to 25% for the background set. The corresponding percentages for CAUUCC within the miR-1 set were 40% and 18%. Also, CD164, which has been predicted to be a miR-124 target on the basis of conservation of potential binding sites among vertebrates³, was the second most significantly downregulated gene in the miR-124 transfection. Significant overlap of conservation-based bioinformatic predictions of miR-1 (refs 3, 19) and miR-124 (ref. 3) targets with the microarray data can be observed at weaker *P* values at the 24 h time point (Supplementary Discussion), implying that our stringent *P* values, even while capturing a large number of probable targets, are still missing others.

We conclude that miR-1 and miR-124, and presumably other tissue-specific or cell type-specific miRNAs, help define and maintain the different cell types of animals. One possibility is that tissue-specific miRNAs may be directly responsible for the lower level of expression of some transcripts within that tissue, similar to how plant miRNAs are thought to help clear cells of unwanted transcription factors during development²⁰. There is precedent for metazoan miRNAs directing cleavage reactions *in vivo*, but so far this has only been shown for target sites with near-perfect complementarity⁹. Although *in vitro* experiments have shown that an RNA-inducing silencing complex (RISC) can cleave highly mismatched substrates^{21,22}, we speculate that the often subtle decreases in transcript levels observed in our transfection experiments are primarily independent of the RISC endonuclease, perhaps arising from the recruitment of other mRNA degradation machinery or as a secondary effect of inhibiting productive translation.

Some observed miRNA targets may already be transcribed at low levels within the tissue in which the miRNA is expressed. Some of these targets might require a level of repression that the transcriptional apparatus does not provide, whereas other 'neutral' targets might have accumulated miRNA complementary sites without

negative consequences. In both of these cases, miRNA target sites might act as a safety mechanism to dampen expression, allowing more flexibility for transcriptional fluctuations on both cellular and evolutionary time scales. Among the downregulated gene sets, we noticed some well-studied paralogues that might require additional repression within a tissue: for instance, synaptogyrin 2 (*SYNGR2*) is downregulated during miR-124 transfection and is a non-neural paralogue of the neural-specific gene synaptogyrin 1 (ref. 23). Likewise, twinfilin-1 (*PTK9*) is downregulated during miR-1 transfection and is a non-muscle paralogue of the muscle-specific gene twinfilin-2 (ref. 24).

It is unlikely that all cell type-specific miRNAs will, upon statistical analysis of gene expression data, produce tissue-specific signals as significantly as miR-1 and miR-124; these two miRNAs, in addition to being highly expressed (Supplementary Fig. 1), occur in cell types that are well represented in tissues that are easily isolated (that is, brain and muscle). For example, we tested a human miRNA sequence expressed in embryonic stem cells; although the 3' UTRs of the downregulated genes were enriched for seed matches, no tissue-specific signal was observed (Supplementary Fig. 5). The lack of a tissue-specific signal was not a surprise, given that our gene-expression reference set did not contain data from human embryonic stem cells.

Our results provide experimental support for the hypothesis that vertebrate miRNA targets are plentiful in number^{3,19,25}. Analysis of the Gene Ontology consortium terms²⁶ of the downregulated genes did not reveal any convincing enrichment for functional classes (data not shown), consistent with the idea that miRNA targets are also diverse in function³. Although quite large, the sizes of these signatures are probably still underestimates of the number of miR-1 and miR-124 targets *in vivo* because many targets may not be expressed in HeLa cells, may exhibit slow kinetics of transcript degradation that are beyond the sensitivity of the microarray platform, or may only be translationally repressed.

Our results indicate that microarrays can be used to detect physiologically relevant miRNA:mRNA interactions in vertebrates, that miRNAs downregulate a greater number of transcripts than previously appreciated, primarily through interaction with 3' UTRs, and that tissue-specific miRNAs enforce broad constraints on a tissue's transcript population, helping to define metazoan cell types. Moreover, given that these experiments were performed in an identical manner to standard siRNA experiments, it is probable that many of the off-target siRNA effects observed previously use a mechanism that is shared with on-target miRNAs. □

Methods

Transfections and microarray analysis

Transfections and microarray hybridizations were performed as described previously³. Briefly, HeLa cells were transfected using Oligofectamine (Invitrogen) in six-well plates with 100 nM RNA duplexes (Dharmacon). Mismatches were introduced into some duplexes to help promote activation of the sense strand^{27,28}. Duplexes were as follows (sense/antisense, 5' to 3' orientation): miR-1, UGGAAUGUAAAGAAGUAUGUAA/ACAUAUCUUUUAUCAUUAUAUA; miR-124, UAAGGCACGCGGUGAAUGCCA/GCAUUCACCGCGUGCCUUAUA; 124mut5-6, UAAGGCACGCGGUGAAUGCCA/GCAUUCACCGCGUGCCUUAUA; 124mut9-10, UAAGGCACCGCGGUGAAUGCCA/GCAUUCACCGCGGUGCCUUAUA; chim1-1/124, UGGAAUGUAAAGAAGUAUGCCA/GCAUUCACCGCGUGCCUUAUA; chim1-124/1, UAAGGCACGCGGUGAAUGUAA/ACAUAUCUUUUAUCAUUAUA. Total RNA was purified using the RNeasy kit (Qiagen), and processed and hybridized as fluorescent label reversed pairs to Agilent microarrays. *P* values were calculated from a previously described error model²⁹.

Sequence analysis

Probe transcripts were associated with LocusLink loci using LocusLink sequence accessions, which were also used to map the loci to unique human Unigene transcripts (Hs.seq.uniq, Unigene 161). Mouse orthologues were assigned using EnsMart (<http://www.ensembl.org>). Following removal of polyA tails, sequences were masked for repeats (A. F. Smit & P. Green, RepeatMasker at <http://ftp.genome.washington.edu/RM/RepeatMasker.html>) and analysed by MEME using the 'zoops' model, setting motif width 'w' incrementally from four to ten. Otherwise, MEME default parameters were used. Sequence logos were constructed using WebLogo (<http://weblogo.berkeley.edu>). Background sets of genes were constructed using LocusLink genes that were represented

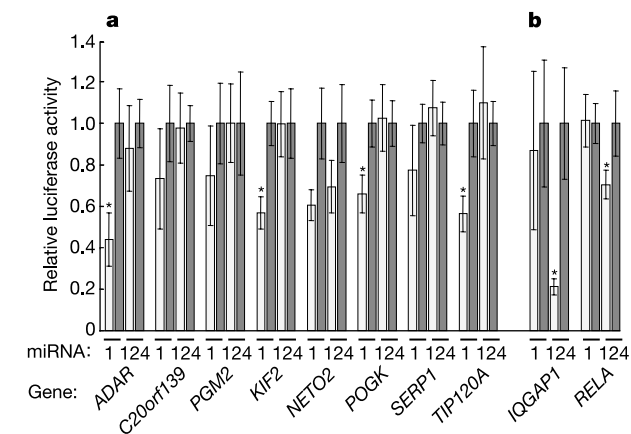


Figure 4 MicroRNA-directed repression of renilla luciferase reporter genes bearing 3' UTR segments from predicted target genes. **a**, **b**, miR-1 (**a**) and miR-124 (**b**) target genes. Duplex siRNAs corresponding to miR-1 and miR-124 were co-transfected with wild-type or mutant reporter plasmids into HeLa cells. After normalizing to the firefly luciferase activity from the transfection control, luciferase values of wild-type (open bars) reporter plasmids were normalized against the average value for the corresponding mutant plasmids (filled bars). The independent normalization of the miR-1 and miR-124 values was required because each of these miRNAs differentially influences the values of the firefly luciferase transfection control. Each mutant plasmid was identical to the wild-type plasmid except for three point substitutions disrupting each of two seed matches. Nine replicates were performed (three independent experiments, each with three culture replicates); error bars indicate one standard deviation. Asterisks denote instances in which miRNA-directed repression was statistically significant when compared with both the mutant reporter and the effect of the non-cognate miRNA ($P < 0.001$; Wilcoxon rank-sum test).

on the chip, and that were either associated with annotated Unigene 3' UTRs or represented in the expression atlas, as appropriate. Orthologous UTRs were aligned with ClustalW³⁰ using default parameters.

Luciferase reporters

Eight miR-1 targets were chosen randomly from those downregulated genes that possessed at least two 6-nt seed matches within a 1-kb segment of their 3' UTR. Two miR-124 targets were also selected, each of which contained two seed matches. Wild-type and mutant UTR segments were cloned into the 3' UTR of either a TK- or SV40-driven renilla luciferase-expressing plasmid, essentially as described³. Transfections were performed as previously⁹, except that 100 nM of miRNA duplex was transfected together with 5 ng of firefly luciferase reporter plasmid (transfection control), 1 µg of pUC19 plasmid (carrier), and either 250 ng of TK-renilla luciferase or 2 ng of SV40-renilla luciferase reporter plasmid. The cloned sequences and putative binding sites are described in the Supplementary Note. Relative luciferase values derived from cognate miRNA with wild-type plasmid cotransfections were compared (by Wilcoxon rank sum test), separately, to values from cognate miRNA with mutant plasmid, non-cognate miRNA with wild-type plasmid and non-cognate miRNA with mutant plasmid cotransfections.

Expression atlas

The human tissue expression atlas was described in the Supplementary Information of ref. 13, where the expression level of each gene is represented by the median intensity of its exon-exon junction probes. Because clustering analysis suggested that the intraventricular septum, pancreas and corpus callosum microarray samples had been mislabelled or mishybridized, these samples were not used in the analysis. Intensities that fell below 200 were not used to assign ranks, because they would be less than a standard deviation above the chip background signal; tissues that fell below this intensity were assigned the same ranks.

Received 22 July; accepted 22 December 2004; doi:10.1038/nature03315.

Published online 30 January 2005.

1. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
2. Carrington, J. C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).
3. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
4. Palatnik, J. F. *et al.* Control of leaf morphogenesis by microRNAs. *Nature* **425**, 257–263 (2003).
5. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotechnol.* **21**, 635–637 (2003).
6. Hutvagner, G. & Zamore, P. D. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**, 2056–2060 (2002).
7. Zeng, Y., Yi, R. & Cullen, B. R. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl Acad. Sci. USA* **100**, 9779–9784 (2003).
8. Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev.* **17**, 438–442 (2003).
9. Yekta, S., Shih, I. H. & Bartel, D. P. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594–596 (2004).
10. Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
11. Sempere, L. F. *et al.* Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.* **5**, R13 (2004).
12. Pruitt, K. D. & Maglott, D. R. RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Res.* **29**, 137–140 (2001).
13. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
14. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
15. Lim, L. P. *et al.* The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* **17**, 991–1008 (2003).
16. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
17. Lai, E. C. Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature Genet.* **30**, 363–364 (2002).
18. Doench, J. G. & Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
19. John, B. *et al.* Human MicroRNA targets. *PLoS Biol.* **2**, e363 (2004).
20. Rhoades, M. W. *et al.* Prediction of plant microRNA targets. *Cell* **110**, 513–520 (2002).
21. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nature Struct. Mol. Biol.* **11**, 599–606 (2004).
22. Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975–980 (2004).
23. Janz, R. & Sudhof, T. C. Cellulogyrin, a novel ubiquitous form of synaptogyrin that is phosphorylated by pp60c-src. *J. Biol. Chem.* **273**, 2851–2857 (1998).
24. Vartiainen, M. K., Sarkkinen, E. M., Matilainen, T., Salminen, M. & Lappalainen, P. Mammals have two twinfilin isoforms whose subcellular localizations and tissue distributions are differentially regulated. *J. Biol. Chem.* **278**, 34347–34355 (2003).
25. Bartel, D. P. & Chen, C. Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nature Rev. Genet.* **5**, 396–400 (2004).
26. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
27. Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199–208 (2003).
28. Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**, 209–216 (2003).

29. Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
30. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements Thanks to S. Baskerville, M. Cleary and P. Sharp for comments on the manuscript, C. Armour, S. Bartz, J. Burchard, G. Cavet, D. Haynor, A. Jackson, M. Pellegrini, E. Schadt and Y. Wang for their assistance, the Rosetta Gene Expression Laboratory for microarray work, M. Jones-Rhoades for primer design, and W. Johnston for plasmid construction.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.P.L. (lee_lim@merck.com). The GEO accession number for the microarray data is GSE2075.

Highly coupled ATP synthesis by F₁-ATPase single molecules

Yannick Rondelez^{1,2}, Guillaume Tresset^{1,2}, Takako Nakashima², Yasuyuki Kato-Yamada³, Hiroyuki Fujita⁴, Shoji Takeuchi⁴ & Hiroyuki Noji²

¹LIMMS/CNRS-IIS, ²Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan

³Department of Life Science, College of Science, Rikkyo (St Paul's) University, Tokyo 171-8501, Japan

⁴CIRMM, Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan

F₁-ATPase is the smallest known rotary motor, and it rotates in an anticlockwise direction as it hydrolyses ATP^{1–3}. Single-molecule experiments^{6–9} point towards three catalytic events per turn, in agreement with the molecular structure of the complex¹⁰. The physiological function of F₁ is ATP synthesis. In the ubiquitous F₀F₁ complex, this energetically uphill reaction is driven by F₀, the partner motor of F₁, which forces the backward (clockwise) rotation of F₁, leading to ATP synthesis^{11–13}. Here, we have devised an experiment combining single-molecule manipulation and microfabrication techniques to measure the yield of this mechanochemical transformation. Single F₁ molecules were enclosed in femtolitre-sized hermetic chambers and rotated in a clockwise direction using magnetic tweezers. When the magnetic field was switched off, the F₁ molecule underwent anticlockwise rotation at a speed proportional to the amount of synthesized ATP. At 10 Hz, the mechanochemical coupling efficiency was low for the $\alpha_3\beta_3\gamma$ subcomplex (F₁^ε), but reached up to 77% after reconstitution with the ϵ -subunit (F₁^{ε+}). We provide here direct evidence that F₁ is designed to tightly couple its catalytic reactions with the mechanical rotation. Our results suggest that the ϵ -subunit has an essential function during ATP synthesis.

Genetically engineered, single F₁ molecules can be grafted on to a glass slide, and the ATP-driven anticlockwise rotation has been observed directly under an optical microscope by coupling the γ -subunit to a fluorescent filament or a submicrometre-sized plastic bead^{9,14}. Our experiment used a similar procedure with some alternations: the magnetic bead was attached to the γ -subunit to rotate it relative to the glass-bound $\alpha_3\beta_3$ stator part of the F₁ molecule using magnetic tweezers¹⁵. We also developed a micro-fabrication technique to confine this system to a 6-fl transparent container¹⁶ (Fig. 1b). Owing to the extremely small volume, the number of ATP molecules produced by a single F₁ enzyme reached a

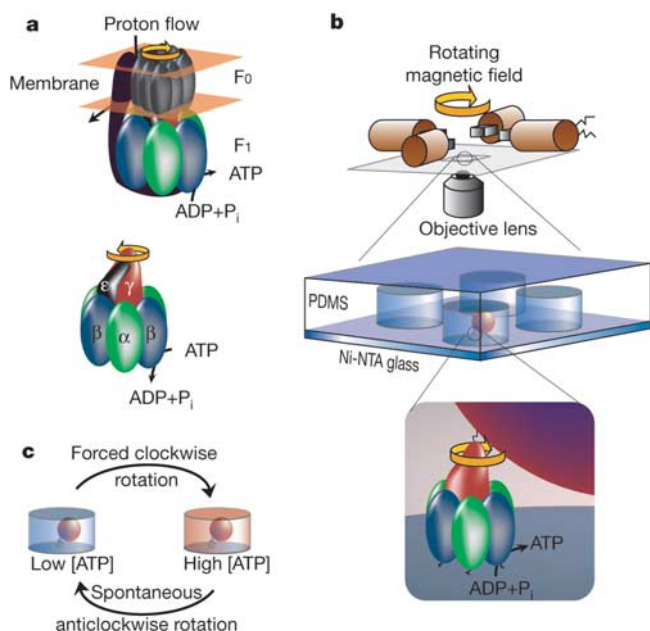


Figure 1 ATP synthesis by F_1 -ATPase. **a**, Schematic view of the membrane-embedded F_0F_1 -ATP synthase in which the proton-driven F_0 rotates F_1 in a clockwise direction for ATP synthesis. Isolated F_1 only hydrolyses ATP, and hence is called F_1 -ATPase. F_1 -ATPase comprises five different subunits, $\alpha_3\beta_3\gamma\delta\epsilon$, of which the $\alpha_3\beta_3\gamma$ subcomplex is the minimal functional part of the ATPase³⁰. ATP hydrolysis occurs sequentially on the three β -subunits, which generate power strokes driving the anticlockwise rotation of $\gamma\epsilon$ (or γ if ϵ is absent). **b**, To force the ATP synthesis of F_1 against the chemical potential, a magnetic bead was attached to the γ -subunit of an immobilized F_1 , and rotated clockwise using magnetic tweezers. A single bead–enzyme complex was enclosed in a PDMS container in order to trap synthesized ATP molecules. **c**, ATP synthesis was detectable via the enzyme itself: after the magnetic field was released, the enzyme resumed its ATP-hydrolysing anticlockwise rotation at a speed proportional to the ATP concentration in the chamber.

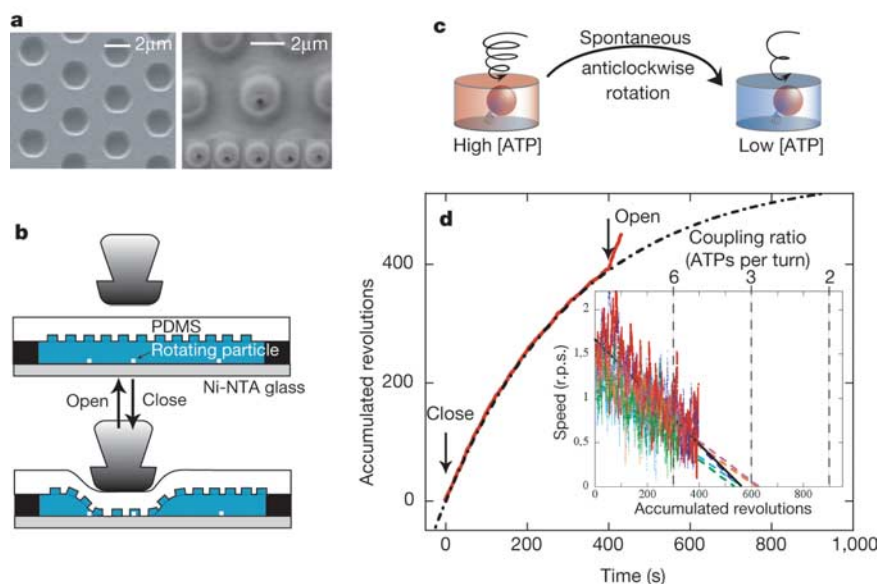


Figure 2 Determination of the coupling efficiency of ATP hydrolysis. **a**, Scanning electron microscopy image of a PDMS sheet with cylindrical cavities of $1.1\text{-}\mu\text{m}$ radius and $1.5\text{-}\mu\text{m}$ height (left), and a bright-field image under an optical microscope after enclosing a rotating bead (right), with successive images of rotation. **b**, Elastic properties of PDMS allow the reversible closing/opening of the chamber, by partially pressing a suspended sheet against the bottom glass plate. **c**, The chamber volume was small enough to ensure that a single enzyme exhausted most of the available ATP (1,800 molecules) within a few

detectable micromolar concentration within minutes of forced clockwise rotation. After the magnetic field was switched off, this ATP concentration could be readily determined by monitoring the speed of the enzyme spontaneously rotating in an anticlockwise direction. By comparing the number of turns in both directions (mechanical input/output) with the changes in ATP concentration (chemical input/output), we were able to assess directly the mechanochemical coupling efficiencies of F_1 .

As a first step, we focused on the enzyme acting as a motor; that is, hydrolysing ATP. F_1 molecules were immobilized on a glass plate at a density of less than one per $4\text{-}\mu\text{m}^2$ (the surface area of a chamber). A polydimethylsiloxane (PDMS) sheet with microchambers on its surface (Fig. 2a, left) was suspended above the glass plate and partially pressed down onto it with a small, flat-tipped glass needle to enclose a rotating F_1 molecule in a chamber (Fig. 2b). The chamber size was large enough to avoid physical conflict with the rotating bead (Fig. 2a, right), and could contain only $1,800 \pm 200$ ATP molecules at 500 nM ATP. (The error number was derived from the error of the chamber volume, around 10%.) Because the motor consumed ATP molecules in the chamber, the rotation speed decreased (Fig. 2c). Under these conditions of low ATP concentration, where ATP binding is the rate-limiting step, the speed decreases linearly with ATP concentration and the expected first-order decay was observed (Fig. 2d). We re-opened the chamber when the speed approached zero, thus taking back the system to its initial condition (Fig. 2b). The enzyme immediately resumed its initial speed, showing that the motor was still active. Extrapolating the rotation speed versus the number of accumulated rotations (Fig. 2d, inset) yielded the maximum number of turns that the single enzyme could perform by exhausting the 1,800 available ATP molecules. We determined the value to be 580 ± 100 turns (mean $\pm$ s.d.), corresponding to a coupling ratio of 3.15 ± 0.5 ATPs per turn. A similar result was obtained for F_1^{+e} (data not shown). Such a high coupling ratio in the motor direction had been proposed previously^{8,17}, but only indirect evidence had been provided so far, because the mechanical output and the real ATP

minutes. **d**, Time course of accumulated revolutions showing an exponential decay after closing the chamber, and the recovery of the initial velocity after opening the chamber. The inset shows the rotation speeds as a function of the accumulated revolutions for six different F_1 molecules. The linear decreases were extrapolated to give the maximum number of possible turns and the coupling ratio between ATP hydrolysis and mechanical rotation.

consumption could not be monitored simultaneously. In particular, the possibility that the motor might 'slip' and hydrolyse ATP without rotation was a lingering question. Back steps⁸ could also affect the coupling ratio; however, this was such a rare event (less than 1% in our case) that it can be considered to be negligible. The present experiment directly demonstrates that F_1^{+e} and F_1^{-e} do not waste ATP, and work as motors with nearly perfect mechanochemical coupling efficiencies.

In a second series of experiments, we turned to the more challenging question of ATP synthesis. The buffer now included physiological concentrations of ADP and inorganic phosphate P_i , as well as 200 nM ATP, in order to identify active rotating molecules. As before, the chamber was closed around a single enzyme, but its rotation was now forced into the clockwise direction using a

rotating magnetic field and the magnetic bead as a handle. The maximum rate of ATP synthesis for this thermophilic F_0F_1 enzyme was biochemically determined to be 30 ATPs s^{-1} at room temperature (R. Iino and M. Yoshida, personal communication). Therefore, we selected a rotating speed of 10 rotations per second (r.p.s.). The magnetic field was applied in repetitive sequences of 1 min duration, with an interval of ~ 1 min (Fig. 3a, b). When it was turned off, the bead resumed its anticlockwise rotation at a speed proportional to the ATP concentration.

If perfect mechanical coupling is assumed, the enzyme should produce three ATPs per clockwise turn. The concentration increment after 10 Hz rotation for 1 min corresponds to 500 nM ATP in 6 fL. This is a significantly higher concentration than the initial ATP concentration of 200 nM, thus one would expect a marked

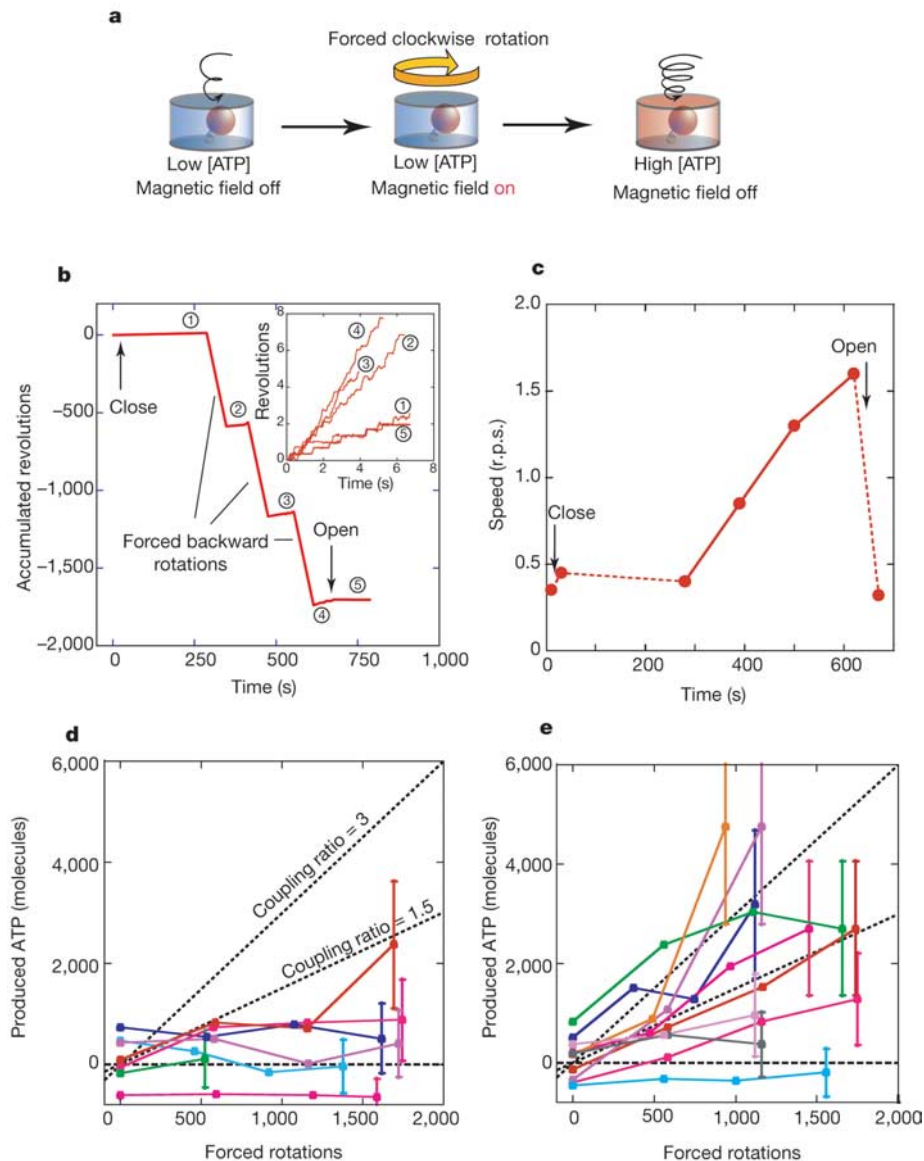


Figure 3 Determination of the coupling efficiency of ATP synthesis. **a, b,** An $F_1^{-e/+e}$ enzyme bound to a magnetic bead was rotated in a backward (clockwise) direction with magnetic tweezers within a microchamber in the presence of ADP, P_i and a small amount of ATP. Synthesized ATP accumulated in the chamber. Before and after a forced clockwise rotation, spontaneous anticlockwise rotation was recorded as a means of detecting ATP synthesis through the increased rotational speed (see inset). **c,** The anticlockwise rotation speeds at each experimental step in **a** are shown. Between the second and third point, the

motor was stalled without rotation in order to confirm the absence of background ATP synthesis. **d, e,** Anticlockwise rotation speeds were converted into the total number of synthesized ATPs for F_1^{-e} (**d**) and F_1^{+e} (**e**). Each trace belongs to an individual F_1 molecule. Dotted lines indicate ATP yields expected for mechanochemical coupling ratios: 3 for 100% efficiency; 1.5 for 50%; and 0 for 0%. The standard error deviation from the rotational velocity determination (30%) and the chamber volume (10%) is indicated at each end of the traces only, for the sake of clarity.

acceleration of the anticlockwise rotation after release of the magnetic field. Notably, this was not the case for F_1^e : the beads hardly increased their speed. Analysis of these experiments (Fig. 3d) gave a coupling ratio of 0.5 ± 0.4 (mean $\pm$ s.d.) ATPs produced per turn. This is in agreement with a previous bulk experiment¹³, indicating a very low mechanochemical coupling efficiency of F_1^e at 10 Hz.

The results were different for F_1^{+e} , reconstituted from F_1^e and the ϵ -subunit. In this case, the forced backward rotations led, in most experiments, to a large increase in the rotation speed (Fig. 3b, c). These reproducible patterns confirmed the production of ATP inside the microchamber. For computation of the ATP yield from the rotation speed, calibration curves of the speed versus ATP concentration (ranging from 200 nM to 2 mM) were determined in independent experiments in the same buffer (see Supplementary Information). It followed Michaelis–Menten kinetics. We then calculated the number of ATP molecules produced from the increment of ATP concentration and the chamber volume to compare with the number of forced clockwise rotations (Fig. 3e). Most traces were found between dotted lines for coupling ratios of 3 and 1.5. We obtained an average value of 2.3 ± 1.6 (mean $\pm$ s.d.) ATPs per turn (77%) after eliminating obviously defective traces (blue and grey traces in Fig. 3e). This is probably a lower limit, because most artefacts, such as transient leaks in the chamber, defective connections in the glass– F_1 –streptavidin–bead complex, or contamination by other free motors, should decrease this value. Therefore our data point to an excellent mechanochemical coupling efficiency. In the best cases, we observed the postulated value of three ATPs synthesized per turn.

Efficiency of molecular machines is generally difficult to assess because it requires precise mechanical information—available only at the single-molecule level—while simultaneously measuring minute amounts of chemical output¹⁷. Moreover, it is easier to detect a mechanical output from a chemical input than to do the contrary. The mechanical synthesis of ATP by reversing F_1 molecules was previously reported¹³, but it was not a quantitative measurement. As a consequence, a good understanding of the ATP-hydrolysing function of F_1 has been accumulated^{1,3,4}, but very little was known about its biologically more relevant ATP synthesis ability². In this approach we have taken advantage of transparent containers so small that a single enzyme's chemical activity quickly results in detectable concentration changes¹⁶. Our experiments show that both F_1^e and F_1^{+e} consume strictly three ATPs per anticlockwise turn, whereas only F_1^{+e} is efficient enough to produce ATP when forced to rotate in the backward (clockwise) direction. These results are consistent with the ubiquity of this strategic enzyme that fuels most of the energy-consuming biological processes.

The present work reveals the unexpected importance of the ϵ -subunit in the synthesis of ATP. Its regulatory role for ATP hydrolysis is well recognized^{18–20} but its exact function in synthesis is still a matter for discussion. Our findings suggest that the ϵ -subunit is necessary for high-yield and high-rate ATP synthesis. It was already known that the absence of the ϵ -subunit suppresses the synthesizing ability of chloroplast F_0F_1 ²¹, even though the enzyme is structurally intact. Recent biochemical studies have suggested that ϵ can insert its helical domain into the $\alpha_3\beta_3$ ring, in a cleft along the γ -subunit^{22–24}. This could be the structural basis for a switch of the enzyme into the ATP synthesis mode²³, and/or a shift in the ATP/ADP affinity balance². Alternatively, as shown in the crystal structure²⁵ of F_1^{+e} , the ϵ -subunit may stabilize the protruding part of the γ -subunit where the mechanical torque is applied. In this case, it would possess mostly a structural function. The precise role of the ϵ -subunit in ATP synthesis remains to be established. The experimental set-up described here can also be used to investigate the parameters of the original mechanical-to-chemical biological transduction performed by F_1 , such as maximum velocity, torque, energy barriers and affinities for substrates. □

Methods

Materials

All buffers were prepared with 0.05 mg ml^{−1} heat-shock-purified BSA, 50 mM MOPS pH 7, 50 mM KCl and 2 mM MgCl₂. For the ATP-driven rotation assay, 500 nM ATP was added, whereas for ATP synthesis the buffer contained 200 nM ATP, 10 mM P_i and 100 μ M ADP, which was purified as previously described¹³. In buffers containing both BSA and ADP, spontaneous ATP synthesis activity was detected owing to contaminating adenylate kinase in BSA. However, the activity was negligible, less than 2 nM h^{−1}. A mutant F_1^e (His₆– α C193S, His₁₀– β and γ S107C/I210C), derived from the thermophilic *Bacillus* PS3, was prepared and biotinylated as described²⁶. Six His tags were used to strengthen the binding of F_1 to the glass plate, but did not alter significantly the rotary and catalytic properties of F_1 as compared to previous mutants⁹. Recombinant ϵ -subunit was prepared as described⁷ and mixed at 7.5 μ M with 0.5 μ M F_1^e , in order to reconstitute F_1^{+e} . PDMS sheets with microchambers were prepared by conventional moulding procedures²⁷ with a slight modification¹⁶. The dimensions of the chambers, determined from electron microscope images, were 1.13 μ m radius and 1.50 μ m height, with an average volume of 6.0 ± 0.6 fl (mean $\pm$ s.d.).

Equipment

The magnetic field was generated by custom-made magnetic tweezers¹⁵ mounted on the stage of a microscope (Olympus, IX71). The bead rotation was recorded using a CCD camera (MTI, RC300) at the video rate, and the trajectories of rotation were analysed using custom-made software (R. Yasuda). During the time intervals when the magnetic field was cut off, the speed of ATP-driven rotation was calculated from more than five successive turns²⁸.

ATP-driven rotation assay in a chamber

A 2×20 mm² flow cell was constructed by suspending a PDMS sheet on top of a Ni-NTA-modified cover plate with two 40- μ m spacer sheets. F_1 molecules were immobilized on the cover plate at a very low concentration (4.7 pM), such that the density of immobilized F_1 was less than one enzyme per 4 μ m² (the surface of a chamber). Streptavidin-coated magnetic beads (Seradyn; $\sim$ 0.47- μ m diameter) were then attached to the biotinylated γ -subunit. When a rotating bead was found, it was enclosed in a microchamber by pressing down on the PDMS sheet using a glass needle (30- μ m diameter) controlled by a z-axis micromanipulator (Narishige). After visual confirmation of contact between the sheet and the cover plate, the glass needle was pressed down a further 2–3 μ m to seal the chambers tightly. This value is less than 10% of the value required for the complete compression (more than 30 μ m). Thus, taking the elastomeric properties of PDMS²⁹ into account, the chamber volume was not strongly modified under these conditions; this is also supported by the ATP-driven rotation experiments. On rare occasions, small focus drifts and transient leakage of the chamber were observed, in which case the entire experiment was rejected.

Mechanical ATP synthesis in a chamber

The set-up was similar to previous experiments, but magnetic tweezers and synthesis buffer were used. Under these high ADP conditions, more than 90% of F_1^e or F_1^{+e} molecules were in the Mg-ADP-inhibited form (data not shown). Therefore, most of the magnetic beads did not show rotation, and they had to be pushed into the active state by mechanical, 180° clockwise rotation¹⁵. A 10 Hz rotating magnetic field was applied for 1 min to force 600 turns in the clockwise direction. Some beads (possibly with a lower magnetic content) did not perfectly follow the rotating field, and thus the actual number of clockwise turns had to be directly counted from the video analysis. Between rounds of the forced rotation, speed of the ATP-driven anticlockwise rotation was measured. The standard curves (see Supplementary Information) were built from independent experiments in regular flow cells. Michaelis–Menten parameters (F_1^e : $K_m = 3.3$ μ M, $V_{max} = 4.5$ r.p.s.; F_1^{+e} : $K_m = 2.1$ μ M, $V_{max} = 4.9$ r.p.s.) and the chamber volumes were used to convert rotational speed into the total number of synthesized ATP molecules.

Received 17 August; accepted 8 December 2004; doi:10.1038/nature03277.

- Boyer, P. D. The ATP synthase—a splendid molecular machine. *Annu. Rev. Biochem.* **66**, 717–749 (1997).
- Senior, A. E., Nadanaciva, S. & Weber, J. The molecular mechanism of ATP synthesis by F_1F_0 -ATP synthase. *Biochim. Biophys. Acta* **1553**, 188–211 (2002).
- Yoshida, M., Muneyuki, E. & Hisabori, T. ATP synthase—a marvellous rotary engine of the cell. *Nature Rev. Mol. Cell Biol.* **2**, 669–677 (2001).
- Kinosita, K. Jr, Yasuda, R. & Noji, H. F_1 -ATPase: a highly efficient rotary ATP machine. *Essays Biochem.* **35**, 3–18 (2000).
- Leslie, A. G. & Walker, J. E. Structural model of F_1 -ATPase and the implications for rotary catalysis. *Phil. Trans. R. Soc. Lond. B* **355**, 465–471 (2000).
- Noji, H., Yasuda, R., Yoshida, M. & Kinosita, K. Jr Direct observation of the rotation of F_1 -ATPase. *Nature* **386**, 299–302 (1997).
- Kato-Yamada, Y., Noji, H., Yasuda, R., Kinosita, K. Jr & Yoshida, M. Direct observation of the rotation of ϵ subunit in F_1 -ATPase. *J. Biol. Chem.* **273**, 19375–19377 (1998).
- Yasuda, R., Noji, H., Kinosita, K. Jr & Yoshida, M. F_1 -ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell* **93**, 1117–1124 (1998).
- Yasuda, R., Noji, H., Yoshida, M., Kinosita, K. Jr & Itoh, H. Resolution of distinct rotational substeps by submillisecond kinetic analysis of F_1 -ATPase. *Nature* **410**, 898–904 (2001).
- Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. Structure at 2.8 Å resolution of F_1 -ATPase from bovine heart mitochondria. *Nature* **370**, 621–628 (1994).
- Zhou, Y., Duncan, T. M. & Cross, R. L. Subunit rotation in *Escherichia coli* F_0F_1 -ATP synthase during oxidative phosphorylation. *Proc. Natl Acad. Sci. USA* **94**, 10583–10587 (1997).

12. Diez, M. *et al.* Proton-powered subunit rotation in single membrane-bound F_0F_1 -ATP synthase. *Nature Struct. Mol. Biol.* **11**, 135–141 (2004).
13. Itoh, H. *et al.* Mechanically driven ATP synthesis by F_1 -ATPase. *Nature* **427**, 465–468 (2004).
14. Hirono-Hara, Y. *et al.* Pause and rotation of F_1 -ATPase during catalysis. *Proc. Natl Acad. Sci. USA* **98**, 13649–13654 (2001).
15. Hirono-Hara, Y. *et al.* Activation of pausing F_1 -motor by external force. *Proc. Natl Acad. Sci. USA* (in the press).
16. Rondelez, Y. *et al.* Microfabricated array of femtoliter chambers allow single molecule enzymology. *Nature Biotechnol.* (in the press).
17. Nishizaka, T. *et al.* Chemomechanical coupling in F_1 -ATPase revealed by simultaneous observation of nucleotide kinetics and rotation. *Nature Struct. Mol. Biol.* **11**, 142–148 (2004).
18. Laget, P. P. & Smith, J. B. Inhibitory properties of endogenous subunit ϵ in the *Escherichia coli* F_1 ATPase. *Arch. Biochem. Biophys.* **197**, 83–89 (1979).
19. Dunn, S. D., Zadorozny, V. D., Tozer, R. G. & Orr, L. E. Epsilon subunit of *Escherichia coli* F_1 -ATPase: effects on affinity for aurovertin and inhibition of product release in unisite ATP hydrolysis. *Biochemistry* **26**, 4488–4493 (1987).
20. Kato, Y. *et al.* Thermophilic F_1 -ATPase is activated without dissociation of an endogenous inhibitor, ϵ subunit. *J. Biol. Chem.* **272**, 24906–24912 (1997).
21. Nowak, K. F., Tabidze, V. & McCarty, R. E. The C-terminal domain of the ϵ subunit of the chloroplast ATP synthase is not required for ATP synthesis. *Biochemistry* **41**, 15130–15134 (2002).
22. Tsunoda, S. P. *et al.* Large conformational changes of the ϵ subunit in the bacterial F_1F_0 ATP synthase provide a ratchet action to regulate this rotary motor enzyme. *Proc. Natl Acad. Sci. USA* **98**, 6560–6564 (2001).
23. Suzuki, T. *et al.* F_0F_1 -ATPase/synthase is geared to the synthesis mode by conformational rearrangement of ϵ subunit in response to proton motive force and ADP/ATP balance. *J. Biol. Chem.* **278**, 46840–46846 (2003).
24. Bulygin, V. V., Duncan, T. M. & Cross, R. L. Rotor/stator interactions of the ϵ subunit in *Escherichia coli* ATP synthase and implications for enzyme regulation. *J. Biol. Chem.* **279**, 35616–35621 (2004).
25. Gibbons, C., Montgomery, M. G., Leslie, A. G. & Walker, J. E. The structure of the central stalk in bovine F_1 -ATPase at 2.4 Å resolution. *Nature Struct. Biol.* **7**, 1055–1061 (2000).
26. Noji, H. *et al.* Purine but not pyrimidine nucleotides support rotation of F_1 -ATPase. *J. Biol. Chem.* **276**, 25480–25486 (2001).
27. McDonald, J. C. & Whitesides, G. M. Poly(dimethylsiloxane) as a material for fabricating microfluidic devices. *Acc. Chem. Res.* **35**, 491–499 (2002).
28. Adachi, K., Noji, H. & Kinoshita, K. Jr Single-molecule imaging of rotation of F_1 -ATPase. *Methods Enzymol.* **361**, 211–227 (2003).
29. Unger, M. A., Chou, H. P., Thorsen, T., Scherer, A. & Quake, S. R. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–116 (2000).
30. Matsui, T. & Yoshida, M. Expression of the wild-type and the Cys-/Trp-less $\alpha_3\beta_3\gamma$ complex of thermophilic F_1 -ATPase in *Escherichia coli*. *Biochim. Biophys. Acta* **1231**, 139–149 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank all members of the Noji and Takeuchi laboratories, and H. Arata and A. Tixier-Mita for discussion and experimental support; R. Yasuda for PC programming of image analysis (CREST image); and Central Workshop in IIS for an optical microscope stage. This work was performed in the framework of LIMMS/CNRS-IIS, and supported in part by Bio-oriented Technology Research Advancement Institution (H.N. and S.T.), and Grants-in-Aid from Ministry of Education, Science, Sports and Culture of Japan (H.N., H.F. and S.T.). Y.R. and G.T. are Research Fellows of the Japan Society for the Promotion of Science.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.N. (hnoji@iis.u-tokyo.ac.jp).

corrigendum

Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution

International Chicken Genome Sequencing Consortium

Nature **432**, 695–716 (2004).

In Table 5 of this Article, the last four values listed in the ‘Copy number’ column were incorrect. These should be: LTR elements, 30,000; DNA transposons, 20,000; simple repeats, 140,000; and satellites, 4,000. These errors do not affect any of the conclusions in our paper. □

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Andy Douglas (4975)
Nils Moeller (4953)
Irene Vigila-Atton (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Advertising Production

Manager: Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:
Tom Hancock

Naturejobs online production:
Niamh Shields

European Satellite Office

**Germany/ Austria/ Italy/
The Netherlands/ Belgium:**

Patrick Phelan
Tel +49 89 54 90 57 11
Fax +49 89 54 90 57 20
e-mail: p.phelan@nature.com
Sharon de Weert
Tel +44 (0) 20 7843 4970
e-mail: s.deveert@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatomachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
Asia-Pacific Sales Director:
Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

Cultivating black gold

A PhD student studying plant biology at the University of Sheffield, UK, has combined his taste for mushrooms, his passion for scientific analysis and his growing interest in business to pitch a plan that has turned him into a funded entrepreneur. Paul Thomas, who is working on light signalling in plants, loves truffles, but they are expensive to buy, difficult to find and almost impossible to cultivate.

But he has used his research background to show in the lab that it could be possible to create the conditions under which truffles grow. The black fungus, whose scarcity makes it worth more by the gram than gold, only appears underground on select tree roots in specific soils and climate. Thomas believes that having a rigorous scientific understanding of these conditions means he could, theoretically, start up a truffle plantation.

But graduate students tend not to have the investment capital needed to snap up land in France or Italy. Nor do they have experience or training in starting small businesses. So Thomas entered a local business-plan competition — something that science graduate students and postdocs are increasingly doing (see *Nature* **428**, 676–677; 2004). Winning the local competition gave him £2,500 (US\$4,600) and the chance to compete for more cash in a BBC reality show called *Dragons' Den*, where would-be entrepreneurs make quick pitches to people who pledge venture capital to the winner (www.bbc.co.uk/dragonsden/episode4.shtml). Thomas was one of the winners there, too, giving up 25% of his company shares for a £75,000 investment.

But Thomas is not about to leave the bench just yet. Truffles take seven years to grow, if all goes right — about the time he needs to become an independent investigator. If his black gold comes up, he may have to pick between the bench and the boardroom of his own company.

Paul Smaglik
Naturejobs editor



Contents

POSTDOCS AND STUDENTS

Finding time for
a family p780

CAREER VIEW

Scientists & Societies
Japan Society for the
Promotion of Science
Graduate Journal
A painful transition
Movers
Harry Finch p782

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Baby blues

Are women being held back by an ‘innate difference’: their ability to give birth? With better academic policies, motherhood and scientific excellence would not seem mutually exclusive, says Virginia Gewin.

Numbers don't lie. Worldwide, PhDs are most often obtained between the ages of 26 and 33. Tack on the series of postdoc positions needed before a professorship is plausible and, for women, the time to have children begins to slip away. For many, conscious that fertility wanes fast after the age of 35, the heavy demands of achieving tenure and motherhood create a tough choice: sacrifice the years spent nurturing a burgeoning career or compromise on plans to raise a family.

The statistics suggest that in the United States most women sacrifice the career. Although women receive 50% of all scientific PhDs in the United States, only 30%, at best, of tenure-track professorships are filled by women. In the United Kingdom, as few as one in 20 professors in science, engineering, technology and maths are women.

CAREER OPTIONS

A department's culture dictates whether a pregnant graduate student will be supported or scorned. For every adviser who demands data delivery from a returning new mother, there is an understanding mentor paving the way.

David Kimelman, a developmental biologist at the University of Washington in Seattle, points out two concerns that come to mind when a student reveals a pregnancy. How will the baby affect the timing and progress of a graduate degree? Will it affect the student's overall interest in the lab?

"The individuals who can become extremely efficient and balance their time are going to be most successful," says Kimelman, noting a trend in recent years towards men playing a more active role in child-rearing. Indeed, one of Kimelman's recent students attributes her effectiveness in the lab to a helpful husband and access to day care nearby.

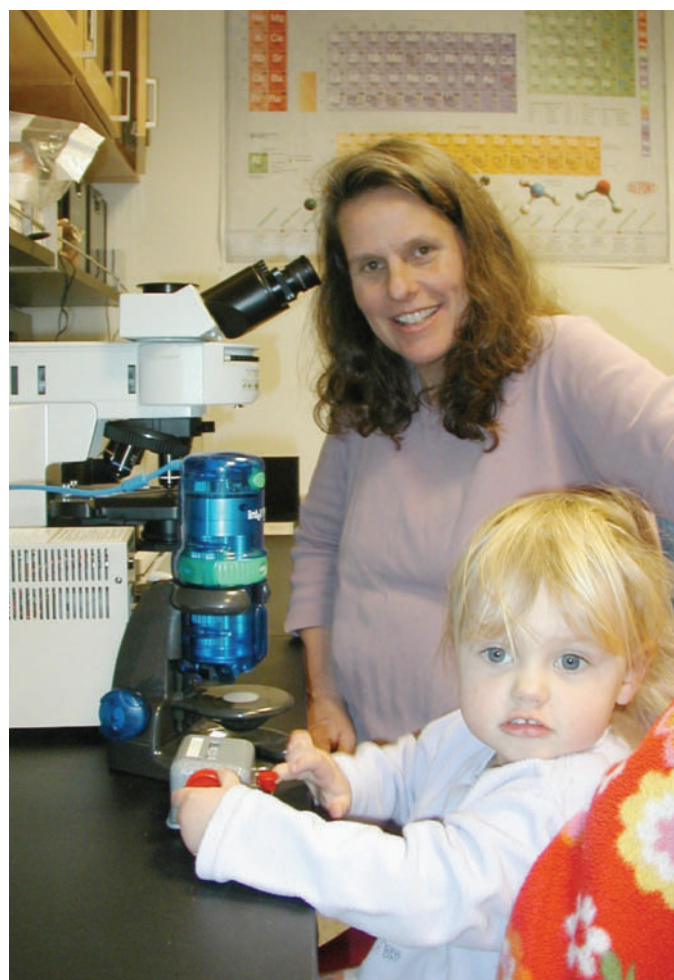
Frank Moore, associate dean of science at Oregon State University, has also noticed the 'daddy day care' trend. His two most recent male doctoral students shared child care responsibilities with their partners.

"This is the good news," says Bob Drago, professor of women's studies at Pennsylvania State University. "For the highly educated in two-parent dual-career relationships, men are doing almost half of the child care," he says. "That's a big improvement that has helped a lot of academic women."

Otherwise, affordable child care is a major stumbling block for most — particularly when good day care options can use up a postdoc stipend. Some

"I'm beginning to accept the possibility that having children may irrevocably alter my career options."

— Heather Leslie



As a tenure-track oceanographer, Margaret Mulholland feels like a pioneer.

campus organizations address the issue. In Britain, the University of York's graduate student association subsidizes nursery care. More often, graduate students use their only other currency — a flexible schedule.

Martha Merrow did just that when she had her children as a graduate student at Tufts Medical School in Boston, Massachusetts. While a postdoc, Merrow, now at the University of Groningen in the Netherlands, sometimes worked "crazy hours" to maximize family time. Up at 5 in the morning to set up experiments, home for breakfast with the family, back to work until dinner and bedtime, then back to work until late in the night.

Thankful for supportive mentors and a stay-at-home husband, the biologist in Merrow points out that she would have had to "protect her genetic investment" if she had been put in a position of making difficult choices early on.

Indeed, early career decisions prove most pivotal, leading to a loss of female researchers before they reach tenure-track stage. "All of our evidence suggests that entering into tenure-track positions is the biggest leak in the academic pipeline," says Mark Goulden, a family-studies researcher at the University of California, Berkeley. He and Mary Ann Mason, dean of Berkeley's graduate division, have shown that women having babies during PhD or postdoc years are less likely to enter tenure-track positions (M. A. Mason and M. Goulden *Academe* 88, 21–27; 2002). Alternatively, if

you look at those in tenure-track positions, men are much more likely than women to be parents.

A 2004 survey of postdocs by Germany's Center of Excellence Women and Science (CEWS) found that almost half of women professors in Germany remain childless, and most felt children would be a burden to their career. Yet there were no differences in scientific productivity between women scientists with and without children.

Not all European countries have such striking differences. For example, Sweden and France tend towards greater equality. CEWS representative Andrea

students have been women for years, she is only the second woman hired as tenure-track member of the 20-odd faculty. "In this day and age it's hard to believe we're still essentially pioneering," she says.

Others, such as UK trade secretary Patricia Hewitt, see the loss of scientifically trained women as a "serious waste of valuable skills, experience and talent". In response to a 2002 study, the Resource Centre for Women in Science, Engineering and Technology opened last autumn to create ways to maximize returns on this human investment. Existing schemes, such as the Daphne Jackson Fellowships, established in 1992, also help women scientists in Britain return to academia.

Many institutions are putting family-friendly policies in place. The European Molecular Biology Organization (EMBO) in Heidelberg, Germany, until recently ran a Restart programme, designed specifically to get women back to the bench. This has now been incorporated into EMBO's general long-term fellowship scheme, allowing women to prolong a 24-month fellowship to 27 months for maternity leave or convert it to part-time for up to three years.

Some 86% of US research institutions allow faculty members to stop the tenure clock for family-related reasons, typically being allowed an extra year to meet department expectations. But, as Kimelman points out, no allowances are made in terms of writing grants or publishing papers. As a result, many don't take advantage of the policy for

fear that it would have a negative effect on their career.

"You can do it all, just not all at once," says Moore. That is the premise guiding the creation of the new Howard Hughes Medical Institute's Janelia Farm research campus in Ashburn, Virginia. Director Gerald Rubin says that Janelia seized the opportunity to create an alternative research environment that eliminates the teaching and grant-writing burdens. "There are only so many hours in a day," says Rubin, adding that the only long-term solution is to adjust the workload from the currently required 80 hours a week. "We're saying spend that extra 30 hours with your family and spend 50 hours on research. Don't put children on hold. Put teaching, administrative functions and grant writing on hold."

Word of this family-friendly policy does not seem to have got out, as a recent call for applications drew only the usual 15–20% of women applicants.

Hogan says that the NSF's Advance programme was designed to challenge institutions to think about how their practices affect women. Since 2001, the agency has awarded more than \$3 million to 19 institutions.

Part-time tenure-track options are also gaining momentum. These arrangements allow a new parent to work half time for half pay and slow the tenure clock accordingly, says Drago. More than 20 US schools, including the major campuses of the University of California system, have adopted such an option so far.

Fortunately for future generations, such proactive schemes take the onus from the individual and put it on the institutions — where it belongs, says Hogan. ■

Virginia Gewin is a freelance science writer in Portland, Oregon.



Family-friendly policies help prevent the choice between family and career.

Löther attributes the difference to better child care options.

Like many, Heather Leslie and her husband Jeremy Rich, postdoctoral fellows at Princeton University in New Jersey, have put off having children until they are more professionally and financially secure. Disheartened by seeing fewer female faces at this stage of her career, Leslie wonders if the increased responsibility and decreased flexibility of a professorship will be stumbling blocks. "I'm beginning to accept the possibility that having children may irrevocably alter my career options," she says.

More senior women also face tough decisions about starting a family while pursuing their career, says Alice Hogan, director of the Advance programme, designed by the US National Science Foundation (NSF) to improve the climate for women in academia. "Later in their lives, some find they did all the things they should have for a successful career and were still marginalized," says Hogan.

ON THE FRINGES

Unfortunately, continued marginalization is a reality. Economist Lawrence Summers, president of Harvard University, caused outrage recently with comments suggesting that "innate differences" play a role in women's under-representation in maths and science.

Margaret Mulholland, an oceanographer at Old Dominion University in Norfolk, Virginia, marvels that, although half of the oceanography graduate

J. CRAIGMYLE/CORBIS

Web links

NSF Advance

♦ www.nsf.gov/home/crssprgm/advance

EMBO Long-Term Fellowship

♦ www.embo.org/fellowships/ltf.html

HHMI Janelia Campus

♦ www.hhmi.org/janelia

Center of Excellence Women and Science

www.cews.org/cews/en

Daphne Jackson Trust

♦ www.daphnejackson.org

UK Resource Centre for Women in Science, Engineering and Technology

♦ www.setwomenresource.org.uk

GRADUATE JOURNAL

A painful transition

Reflecting on my entry from last month (see *Nature* 433, 338; 2005), I am struck with intense pain and a feeling of naivety. When I wrote that, my future was geographically constrained because my wife was my highest priority, my outlook for 2005 was one of hope, commitment, love and anticipation.

After my wife announced she was going to leave me — somewhere between me writing my entry and its publication — my feelings shifted to loss, desperation, anger, frustration and self-loathing. It turns out that we had different perspectives about what my upcoming graduation meant. I saw this crossroads as the end of a difficult period of work. She saw this milestone as an end.

Needless to say, my roller-coaster of emotions has distracted me from my work. I initially fled the lab to deal with my feelings, but now I'm back and trying to focus on getting myself healthy, both emotionally and physically, and finishing my PhD. I am also working hard to address my spite and anger so that we can move forward not as partners, but as close friends. I try to keep in mind the wisdom of Robert Frost: "In three words, I can sum up everything I've learned about life. It goes on." One aspect of my previous journal entry remains true. It is going to be an interesting year. ■

Jason Underwood is a graduate student in microbiology at the University of California, Los Angeles.

SCIENTISTS & SOCIETIES

Japan Society for the Promotion of Science

"The typhoon is advancing — you can finish in one hour," says my *senzai* (assistant professor) and leaves the lab. I am in Japan, well known as 'the land of the rising sun' — although as a visiting fellow from Austria, I have learned that it is also the land of typhoons and earthquakes.

Why did I choose Japan? Because I longed for something different and thought that I could find my personal challenge in Asia. Western scientists working in Japan are relatively rare. In 2003, the Japan Society for the Promotion of Science (JSPS) brought in 144 European researchers and 117 US researchers for short stays, and only 32 Europeans and 15 US scientists for longer-term stints.

I applied for a JSPS short-term stipend and

eventually got an invitation to work for two months at the pharmaceutical sciences department of the University of Osaka. Life as a scientist here is very different from anywhere else I have worked. The Japanese system places a premium on tight supervision, close collaborative effort — and long hours.

Apart from needing permission from my *senzai* to take cover in the event of a typhoon, I also need his permission to begin experiments. Every new experimental idea must be checked by the *senzai* and then the professor. On Fridays, students and junior researchers present a summary of their week's activities and deliver exact research plans for the following week.

A magnetic board monitors the presence and location of everyone in the lab. Right at the beginning, I received my personal marker to indicate exactly where I spent my time. And

Japanese students have to dispose of their rubbish and clean their labs themselves.

Working days begin at 9:30 a.m. and usually end at about 10 p.m., with seminars (in Japanese only) sometimes held on Saturdays. Most students extend their work to weekends and holidays. Sleep deprivation often takes its toll in seminars or in front of the computer — but not to worry, there is no shame in dozing off at any time or anywhere.

Apart from experiencing a different research environment, doing a working visit to this country has allowed me to experience an entirely different way of life. I'd advise Western scientists to consider a visit. But take some language lessons first — and be prepared to work in a more controlled environment. ■

Marika Willeroeder recently completed her PhD in genetics at the University of Salzburg, Austria.

♦ www.jsps.go.jp

MOVERS

Harry Finch, senior vice-president of therapeutics, Argenta Discovery, Harlow, UK



For someone who co-invented an asthma drug that is, so far, worth nearly US\$1.8 billion, Harry Finch didn't exactly make a flying start. He left school at 16 to work in a tar distillery, with just five exam passes. His only science subject was physics-with-chemistry; he wasn't considered good enough at chemistry to take a whole exam in it.

Part-time study followed, and a

string of awards. Everything fell into place when a far-sighted teacher steered him towards organic chemistry.

"I was lucky because, then, I knew I'd found my niche. Very few people find their true niche, the work that matches what they're best at," says Finch.

Married with children by the time he gained his PhD at 27, he opted for a steady job. He found one at Glaxo, where he learned biology from colleagues. "I loved it," says Finch. "And I needed it for drug discovery."

Serendipity played a role in the invention of the beta-agonist salmeterol, he says, quoting Louis Pasteur: "Chance favours the prepared mind." He and his team were working on a cardiovascular drug, but kept noticing that their molecule had the activity sought by the asthma-drug team.

"Finally, we went to them and said: 'Why don't you make this analogue?'," Finch recalls. "They did, and they only

had to make two more changes and that was salmeterol."

When a merger scattered his 280-strong chemistry group in late 2000, Finch switched to biotech companies. His latest move, to Argenta Discovery, has returned him to a familiar area, as the firm works on respiratory therapeutics.

Having experienced both big drug companies and the "white knuckle" world of small biotech, Finch is aware of the advantages of each. In a small biotech you multitask across scientific and discipline boundaries, he says. But a big company provides more people in more disciplines from which to learn.

"I'd encourage people to get experience in at least two places, medium and small or big and small," he says. "Try them both as early as possible and get as much variety of experience as you can." That mix of experience can help young scientists find their niche — just as Finch did after his late start. ■

CV

2004: Senior vice-president, therapeutics, Argenta Discovery, Harlow, UK.

2003–04: Research director, British

Biotech/Vernalis, Cambridge, UK.

2001–03: Research director, RiboTargets, Cambridge, UK.

1978–2001: Research manager to director of chemistry, Glaxo/GlaxoWellcome, UK.

1976–78: Team leader, medicinal chemistry, Allen & Hanbury's, Ware, UK.

1975–76: Senior chemist, Roche Products, Welwyn Garden City, UK.

1972–75: MSc and PhD, University of Manchester, UK.

Undead again

How sweet the taste of freedom.

Ken MacLeod

It's 2045 and I'm still a vampire. Damn.

The chap from Alcor UK is droning through his orientation lecture. New age of enlightenment, new industrial revolution, many changes, take some time to adjust, blah blah blah. I'm only half-listening, being too busy shifting my foot to keep it out of the beam of direct sunlight creeping across the floor, and trying not to look at his neck.

I feel like saying: I've only been dead 40 years, for Chr ... for crying out loud. I saw the first age of enlightenment. I worked nights right through the original Industrial Revolution. I remember being naive enough to get excited about mesmerism, galvanism, spiritualism, socialism, Röntgen rays, rationalism, radium, mendelism, Marconi, relativity, feminism, the Russian Revolution, the bomb, nightclubs, feminism (again), Apollo 11, socialism (again), the fall of Saigon and the fall of the Wall.

The last dodgy nostrum I fell for was cryonics.

So don't give me this futureshock shit, sunshine. The most disconcerting thing I've come across so far in 2045 is the latest ladies' fashion: the old sleeveless mini-dress. The ozone hole has been fixed, and folk are frolicking in the sun. I hug myself with bare arms, and slide the castored chair back another inch.

Under the heel of my left wrist, I feel the thud of my regenerated heart. It beats time to the artery visible under the tanned skin of the resurrection man's neck. The rest of my nature is unregenerate. I feel somewhat thwarted. This is not, this is definitely not, what I died for. And it seemed such a good idea at the time.

It always does.

By 1995 we thought we had a handle on the thing. It's a virus. In all respects but one, it's benign: it prevents ageing and stimulates regeneration of any tissue damage short of, well, a stake through the heart. But it has a very low infectivity, so it takes a lot of mingling of fluids to spread. Natural selection has worked that one hard. Hence the unfortunate impulses. And by 1995, I can tell you, I was getting pretty sick of them. I cashed in my six Scottish Widows life insurance policies (let's draw a veil over how I acquired them), signed up for cryonic preservation in the event of my death, and after a discreet ten years, met an unfortunate and bloody end at the hand of the coven senior, Kelvin.

"You'll thank me later," he said, just before he pushed home the point.



"See you in the future," I croaked.

The last thing I saw was his grin. That, and the pavement below the spiked railings beside the steps of my flat. A tragic accident. The coroner, I just learned, blamed it on the long skirt. Vampires — always the fashion victims.

I leave the orientation room, hang around until dusk under the pretext of catching up with the news, and go out and find a vintage clothes shop. I walk out in Victorian widow's weeds. They fit so well I suspect they were once mine.

"It didn't work," I tell Kelvin.

He sips his bloody mary and looks defensive. "It did in a way," he says. "There are no viruses in your blood."

That word again. I look away. We're in some kind of goth club, which covers for the mode but doesn't improve my mood.

"So why do I still feel ... hungry?"

"Have a tapas," he says. "But seriously ... the way we figure it, the virus has to have transcribed itself into our DNA. So the nanotech cell-repair just replicates it without a second thought."

"So we're stuck with it," I say. "Living in the dark and every so often..."

"Not quite," he says. "Now it's been estab-

lished that cryonics really does work, there's been a whole new interest in a very old idea..."

The coffin lid opens. Kelvin's looking down, as I expect. The real shock is the light, full-spectrum and warm. It feels like something my skin has missed for centuries. I sit up, naked, and bask for a moment.

The overhead lights reproduce the spectrum of Alpha Centauri, which is where we're going. The whole coven is here, all 13 of them, happier and better fed than I've ever seen them. It's taken us a lot of planning, a lot of money, and a lot of lying to get here, but we're on our way.

"Welcome back," says Kelvin. He grins around at the coven.

"Let's thaw one out for her," he says. "She must be hungry."

As far as I can see stretch rows and rows of cryonic coffins containing interstellar colonists in what they euphemistically call cold sleep. Thousands of them.

Enough to keep us going until we reach that kinder sun.

Ken MacLeod's next book is Learning the World. His latest, Newton's Wake, is just out in Orbit paperback. He lives in West Lothian, Scotland.